Министерство образования и науки Российской Федерации
Псковский государственный университет

Л.В. Мотайленко

ОСНОВЫ МОДЕЛИРОВАНИЯ

Учебное пособие

Рекомендовано к изданию кафедрами
«Информационные системы и технологии»
и
«Вычислительная техника»
Псковского государственного университета

Псков
Псковский государственный университет
2013

УДК 681.51
ББК 22.18
М85

Рекомендовано к изданию кафедрами «Информационные системы и технологии» и «Вычислительная техника» Псковского государственного университета

Рецензенты:
- Ильин С.Н., заместитель генерального директора ОАО «СКБ Вычислительной техники»;
- Лёхин С.Н., к.т.н., декан факультета Информатики, заведующий кафедрой «Вычислительная техника»

М85 Мотайленко, Л. В.
Основы моделирования учебное пособие:. — Псков: Изд-во Псков ГУ, 2013. — 104 с.

Учебное пособие предназначено для студентов всех форм обучения факультета Информатики Псковского государственного университета по направлению 230400 «Информационные системы и технологии». Может быть использовано студентами других вузов технических специальностей и направлений подготовки дипломированных специалистов.

Учебное пособие содержит основные понятия и определения теории моделирования, вопросы самоконтроля по дисциплине «Основы моделирования».

УДК 681.51
ББК 22.18

Оглавление

1. Моделирование

1.1. Основные понятия и определения

Моделирование является одной из составных частей аппарата системотехники [1, 3, 11].

Все то, на что направлена человеческая деятельность, называется ***объектом***.

Модель — это система, имеющая некоторые идентичные свойства с оригиналом.

Замещение реального объекта моделью с целью получения информации о важнейших свойствах его называется ***моделированием***. Получение информации о реальном объекте осуществляется путем проведения экспериментов с его моделью на ЭВМ, АВМ, ЭВМ-АВМ или непосредственно на реальном объекте.

Если эксперименты проводятся на ЭВМ, то их называют ***машинными экспериментами***, а сам процесс проведения экспериментов ***имитационным моделированием***. Эксперименты, которые проводятся непосредственно на реальном объекте, называются ***натурными экспериментами***. Эксперименты называются полунатурными, если объект, над которым проводиться эксперимент помещается в моделируемую или искусственную среду, например, при создании новых самолетов эксперименты проводят с помощью аэродинамических труб; плазма, полученная в лабораторной установке, может быть моделью при изучении звездного вещества.

Процесс моделирования можно представить в виде цепочки:

«Описательная модель —математическая схема — математическая модель».

Описательная модель формируется на естественном языке. Под понятием «*математическая схема*» понимают переход от словесного описания системы к формальному представлению процесса и его функционирования в виде некоторой *математической модели* (аналитической или имитационной). Такой подход позволяет рассматривать математику не как метод расчета, а как метод мышления.

При моделировании должны быть рассмотрены три принципа — *концепции модели*:

- рационализм — аналитические методы на базе фундаментальных наук (математика, физика и т.п.);
- эмпиризм (практика) — экспериментально-статистические методы;
- прагматизм — целенаправленный деловой подход, в бизнесе — инженерный, экономический.

Рассмотрим **основные этапы разработки моделей**.

1. Постановка задачи: необходимость решения самой задачи и

моделирования системы; выбор метода решения и определение подзадач.

2. Анализ задачи моделирования системы: выбор критериев эффективности процесса функционирования системы; выбор возможных методов идентификации; предварительный анализ алгоритма модели и результатов моделирования.
3. Определение требований к исходной информации об объекте моделирования и организация её сбора.
4. Выдвижение гипотез и принятие предположений.
5. Определение параметров и переменных модели.
6. Установление основного содержания модели: на этом этапе выбирается метод построения модели.
7. Обоснование критериев оценки эффективности системы.
8. Определение процедур аппроксимации реальных процессов.
9. Описание концептуальной модели.
10. Проверка достоверности концептуальной модели.
11. Составление технической документации.

1.2 Общая классификация моделей систем

Создаваемые модели систем классифицируются по следующим основным признакам:

- *по времени*:
  - динамические модели: непрерывные, которые описываются дифференциальными уравнениями; дискретно–непрерывные (разностные), описываются разностными уравнениями; вероятностные, построенные на событиях — модели теории массового обслуживания;
  - дискретные модели — автоматы;
- *по признаку случайности*:
  - детерминированные — модели, отражающие процессы, в которых явно отсутствуют либо не учитываются всякие случайные воздействия;
  - статистические — модели, отражающие процессы, в которых учитываются случайные воздействия;
- *по назначению*: экономические, технические, военные, социальные, биологические, технологические и т.п.
- *по виду обрабатываемой информации*:
  - информационные: справочно-информационные; информационно - советующие; экспертные; автоматические;
  - физические модели: натурные (плазма); полунатурные (аэродинамические трубы);
  - имитационные модели;
  - интеллектуальные модели;

- семантические (логические) модели и т.п.

1.3 Постановка задачи

Основными целями моделирования объектов является

1) изучение его внутренних свойств;

2) изучение поведения объекта при изменении внешних условий влияющих на него.

Изучение внутренних свойств объекта сопровождается уточнением модели объекта и проверкой адекватности модели объекту. При этом необходим достаточно большой объем априорных (доопытных) данных об объекте, которые получают путем наблюдений за его поведением, проведением натурных испытаний и т.п.

Изучение поведения объекта при изменении внешних условий влияющих на него проводят на базе модели, не требующей изменения в процессе моделирования, и, чаще, данная цель моделирования ставится для изучения поведения проектируемых объектов: новых приборов, схем, когда получены математические выражения, описывающие свойства будущего объекта.

Моделирование возможно и при объединении обеих целей.

При достижении целей моделирования проводится идентификация объекта моделирования. Существует два типа идентификации.

Структурная идентификация — определение внутренней структуры объекта моделирования. Структурная идентификация проводится на базе теории системного анализа.

Параметрическая идентификация **—** определение параметров модели объекта. Проводится на базе математических моделей.

Модель объекта представляет собой черный ящик, которая описывается зависимостью (функцией) выходного сигнала y или η от входного сигнала (рис. 1.1). ε — случайные воздействия на объект, которые могут учитываться в модели.

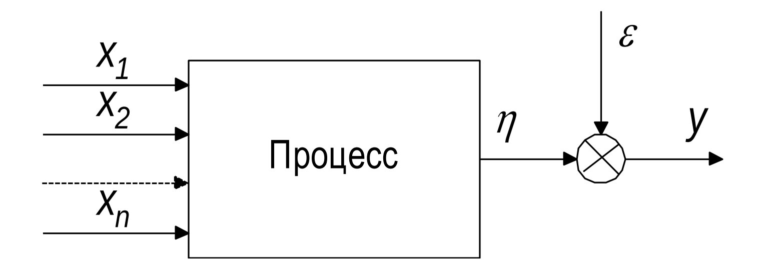


Рисунок 1.1 — Статистическая модель

Количество входных сигналов не ограничено и зависит от

математической сложности функции и учитывающихся в модели факторов. Входы могут называться факторами, входными сигналами, воздействующими сигналами, сигналами воздействия.

Независимые входные сигналы x_i в эксперименте называют ***факторами***. В модель необходимо включить все существующие факторы, которые оказывают влияние на выходную величину η.

Для отсеивания несущественных факторов используют *метод случайного баланса*.

Факторы разделяются на *количественные* и *качественные*. К качественным факторам относятся разные вещества, разные технологические способы, аппараты, исполнители и т.д.

При проведении эксперимента к факторам предъявляются следующие требования:

- факторы должны быть управляемы, т.е. они могут быть установлены на определенный уровень и поддерживать уровень постоянным в течение всего опыта, в случае активного эксперимента. Если уровень фактора изменяется («плывет»), то следует использовать специальный метод, например, замер факторов по возможности должен быть более точным; степень точности устанавливается в зависимости от диапазона изменения факторов;
- факторы должны быть однозначными.

Количество выходов черного ящика — один, и может носить названия: выходной сигнал, функция отклика, реакция модели и, просто, выход.

К выходному сигналу предъявляются следующие требования.

1. Эта величина должна количественно оцениваться, т.е., например, поддаваться измерению. Если нет возможности количественно измерить его, то используют ранговый подход так называемое ***ранжирование***. Величине η присваивают оценки, называемые ***рангами*** по заранее выбранной схеме, например двухбалльная, десятибалльная и т.п. В простейшем случае область определения η может содержать два значения: да и нет, хорошо и плохо (годная продукция – брак).
2. Должна выполняться однозначность в статическом смысле. Это означает, что заданному набору значений x_i должно соответствовать одно с точностью до ошибки экспериментальное значение η. Если ошибка значительная, то необходимо повторять опыты.
3. Должна оценивать эффективность функционирования системы.
4. Желательно, чтобы выходная величина обладала свойством универсальности и полноты, т.е. всесторонне характеризовала объект;

чтобы параметр η имел физический смысл, был простым и поддавался вычислению.

1.4 Стратегия эксперимента

Экспериментально-статистические методы в основном базируются на использовании *пассивного* и *активного эксперимента*.

При ***пассивном эксперименте*** исследователь находится в роли пассивного наблюдателя. Эксперимент ведет сама природа. Экспериментатору приходится только фиксировать значения входных и выходных величин. Модели, полученные методом пассивного эксперимента почти не удается проверить на адекватность.

При ***активном эксперименте*** исследователь вмешивается в процесс эксперимента путем варьирования уровней входных величин.

В рамках активного эксперимента построение модели проходит следующие этапы:

1) выбирается форма модели процесса;
2) строится план эксперимента;
3) проводится экспериментирование;
4) дается анализ результатов эксперимента.

На практике экспериментатору приходится чаще планировать не один, а несколько экспериментов, выполняя и анализируя каждый и, в соответствии с результатами, изменять план эксперимента. Стратегия такого эксперимента показана на рисунке 1.2.

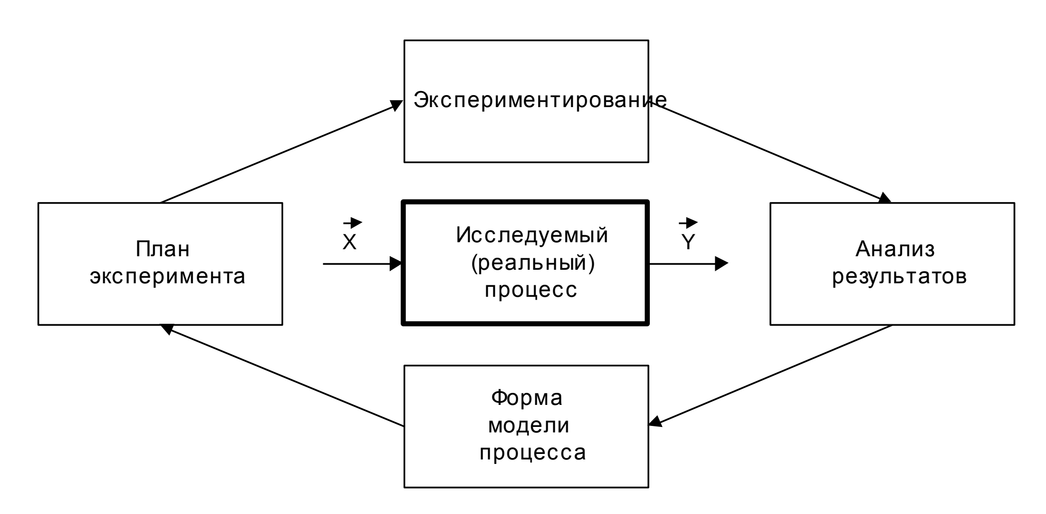


Рисунок 1.2 — Стратегия эксперимента

В результате анализа результатов эксперимента может возникнуть необходимость исправления формы модели и плана эксперимента, тогда эксперименты повторяются вновь по указанной схеме.

1.4 Тактика эксперимента

Стратегическое планирование применяется при любом виде экспериментов, как при натурных, так и машинных экспериментах. При машинных экспериментах дополнительно возникает необходимость обеспечения точности и достоверности результатов, а также количества прогонов.

В связи с этим при тактическом планировании решаются следующие вопросы:

1) определяются начальные условия и их влияние на достижение результатов процессов моделирования;
2) обеспечения точности и достоверности результатов;
3) уменьшение дисперсии оценок характеристики процесса функционирования систем;
4) выбора правил автоматической остановки имитационного эксперимента.

Рассмотрим каждый вопрос подробнее.

Определение начальных условий усложняется тем, что имитационная модель работает эпизодически, т.е. только когда экспериментатор проводит эксперименты. Поэтому, начальный период работы искажается из-за влияния начальных условий запуска машины. Для решения этой проблемы исключают из-за рассмотрения информацию о модели, полученной в предыдущей части моделирования, либо начальные условия выбирают так, чтобы сократить время достижения установившегося режима. Эти приемы позволяют только уменьшить, но не свести к нулю время переходного процесса при проведении машинного эксперимента.

Вопросы обеспечения точности и достоверности результатов, уменьшение дисперсии оценок характеристики процесса функционирования систем и выбора правил автоматической остановки имитационного эксперимента будут рассмотрены в разделах 5.3, 5.4, 4.5 соответственно.

Контрольные вопросы

1. Сформулируйте определение понятия «Моделирование».
2. Сформулируйте определение понятия «Модель».
3. Приведите общую классификацию моделей систем.
4. Идентификация — это?
5. Что такое фактор?
6. Какие требования предъявляются к выходному сигналу?
7. Сформулируйте понятие пассивного эксперимента.
8. Сформулируйте понятие активный эксперимент
9. Перечислите этапы построения модели.
10. Какие вопросы решаются при тактическом планировании.

2. Форма модели процесса

2.1 Имитационное моделирование. Моделирующий алгоритм

Для моделирования процесса на ЭВМ необходимо преобразовать математическую модель в специальный моделирующий алгоритм [8, 12, 14].

Существуют различные способы представления моделирующего алгоритма.

Исторически первым из таких способов является запись алгоритмов при помощи операторных схем. Затем появились языки программирования, пакеты прикладных программ и специальные языки моделирования.

Среди важнейших операторов моделирующих алгоритмов выделим следующие.

1. Вычислительные операторы. Это в основном арифметические операторы. В операторных схемах моделирующих алгоритмов вычислительный оператор может описывать любую сколь угодно сложную и громоздкую группу операций.

2. Операторы формирования реализаций случайных чисел, величин, событий и процессов (см. раздел 4.2, 4.3).

Для имитации действия различных случайных факторов, сопровождающих исследуемый процесс, при моделировании возникает необходимость формировать реализации случайных событий, случайных величин и случайных функций.

Для реализации всех этих перечисленных случайных групп необходимо иметь случайные числа. Операторы формирования реализаций случайных процессов решают задачу преобразования случайных чисел стандартного вида в реализации случайных процессов с заданными свойствами.

3. Операторы формирования неслучайных величин. Они нужны при моделировании детерминированных математических моделей для реализации различных констант и неслучайных функций времени.

2.1.1 Принципы построения моделирующих алгоритмов

В большинстве имитационных моделей имитируется поведение системы на некотором отрезке времени. Поэтому при создании модели и выборе языка программирования важной задачей является определение механизма системного времени. Это необходимо для корректирования временной координаты состояния системы и для обеспечения согласованности различных блоков и событий в системе.

Модель функционирует в искусственном времени и необходимо обеспечить появление событий в определенном порядке с надлежащими временными интервалами между ними.

2.1.2 Принцип фиксированного шага. Принцип переменного шага

Существует два основных метода задания времени: с помощью фиксированных и переменных интервалов времени. Их иногда еще называют метод фиксированного шага Δt и шага до следующего события. По методу фиксированного временного шага отсчет системного времени ведется через заранее определенные временные интервалы. При использовании метода переменного шага состояние моделируемой системы обновляется с появлением следующего события.

В непрерывных моделях используются механизмы фиксированных приращений временных интервалов. В большинстве моделей с дискретным изменением событий используется метод отсчета времени до следующего события.

2.1.3 Принцип особых состояний

Бывают случаи, когда соотношения математических моделей систем удается преобразовать таким образом, что можно определить последующее особое состояние системы по предыдущему особому состоянию или нескольким предыдущим состояниям. Для таких случаев моделирующий алгоритм может быть построен по так называемому принципу "особых состояний". Он отличается от принципа Δt только тем, что включает в себя процедуру определения момента времени, соответствующего следующему особому состоянию по известным характеристикам данного или предыдущего состояния.

2.1.4 Принцип последовательной проводки заявок

В системах массового обслуживания при моделировании процессов обработки заявок иногда удобно строить моделирующие алгоритмы по принципу воспроизведения истории отдельных заявок в порядке поступления их в систему. Алгоритм обращается к сведениям о других заявках лишь в том случае, если это необходимо для решения вопроса о дальнейшем порядке обслуживания данной заявки. Такие алгоритмы весьма экономны, однако они требуют весьма сложную логическую структуру. При моделировании систем массового обслуживания и дискретных производственных процессов, как правило, используют этот принцип, называемый иногда "принцип последовательной проводки заявок". Данный принцип, иногда в сочетании с принципом Δt , лежит в основе почти всех языков моделирования.

2.2 Языки моделирования

Для имитационного моделирования используются как универсальные, так и специальные языки [2,4]. Универсальные языки дают большие возможности программисту в смысле гибкости

разработки, отладки и использования модели. Однако это требует больших усилий, затрачиваемых на программирование. В этом случае усложняется порядок выполнения операций, отсчет системного времени и контроль.

Специализированный язык отличается от универсального своими специфическими свойствами.

К этим свойствам или требованиям можно отнести:

- способность генерировать случайные числа;
- возможность генерировать случайные величины и процессы;
- возможность "продвигать" время либо на одну единицу Δt, либо до следующего события;
- способность накапливать выходные данные;
- способность проводить статистический анализ накапливаемых данных;
- способность распределять выходные данные по заранее заданным форматам;
- способность выполнять идентификацию конкретных событий.

Некоторые из языков имитационного моделирования являются и описательными языками. Они близки к естественному языку, поэтому имитационные модели, написанные на таком языке, легче воспринимаются руководителями и специалистами, не имеющими непосредственного отношения к программированию.

Специальные языки цифрового имитационного моделирования делятся на две группы, соответствующие двум видам имитации: дискретных и непрерывных процессов.

Для моделирования непрерывных процессов используются языки:

- Динамо — для аппроксимации непрерывных процессов используют дифференциальные уравнения первого порядка;
- CSMP, Midas и др., которые помимо блочного построения, применяемого в языке Динамо, еще обладают мощью и удобством алгебраической и логической алгоритмизацией.

Языки для моделирования дискретных процессов можно разбить на четыре категории:

1) языки, ориентированные на действия (CSL, Форсим IV и др.);
2) языки, ориентированные на события (Симскрипт, Симком и др.);
3) языки, ориентированные на процессы (Симула, SOL);
4) языки, ориентированные на потоки сообщений (GPSS, BOSS).

В первой группе действия представлены в модельном времени как мгновенные. В этих языках нет регламентации действиям. Вместо этого исполнительные программы просматривают набор всех условий, от которых зависит появление какого-либо события. И только тогда, когда выполняются все контролируемые условия, происходит изменение состояния и сдвиг времени в этой части программы.

Некоторые задачи удобнее программировать на языке ориентированном на события. При этом событие регламентировано. Регламентация обеспечивает наступление события именно в тот момент времени, когда динамическое состояние показывает, что сложились условия для его появления.

Языки, ориентированные на процессы, объединяют достоинства первых двух языков, т.е. краткость языков, ориентированных на действия и эффективность языков, ориентированных на события. Написанная на этом языке программа работает как несколько независимых программ: одна посредством просмотра действий, другая посредством регламентирования событий.

Что касается четвертой группы языков, то они на самом деле являются языками процессов и отличаются только схемным построением.

Язык GPSS обычно выделяют в отдельную группу. Язык GPSS представляет собой интерпретирующую языковую систему, которую применяют для описания пространственного движения объектов.

2.3 Виды моделей

Стратегия эффективного планирования эксперимента начинается с выбора формы модели процесса [7]. В зависимости от априорной информации о процессе различают две задачи:

- построение модели при неизвестной структуре процесса;
- нахождение параметров модели при заданной структуре.

Априорная информация о процессе для первого случая почти отсутствует. При этом необходимо решать задачу выбора структуры модели процесса, устанавливать класс моделей, оценивать степень стационарности и линейности процесса и др. Решение подобных задач связано с большими трудностями. На заданном этапе приходится решать вопросы содержательного неформального эвристического характера.

При известной функциональной зависимости между выходными и входными величинами задача является более простой и доступной. При этом форма модели известна. Экспериментатору остается построить план, провести эксперимент и сделать анализ результатов экспериментов.

2.3.1 Модели в статике

Модели в статике бывают следующего вида (рис. 2.1).



Рисунок 2.1 — Схема модели в статике

1. Линейная по входным величинам x и по параметрам:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (2.1)$$

2. Нелинейная по x и линейная по коэффициенту β:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \ldots + \beta_n x_n^n$$

3. Линейная по x и нелинейная по коэффициенту β:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2^2 x_2 + \ldots + \beta_n^n x_n$$

4. Нелинейная по x и по β:

$$\left.\begin{aligned} \eta &= e^{\beta_1 x_1} + e^{\beta_2 x_2} + \ldots + e^{\beta_n x_n} \\ \eta &= \sqrt{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n} \end{aligned}\right\},$$

где η — реакция модели на входные величины x.

Встречаются модели и другого вида, например, тригонометрические, логические и др.

Определение функциональной связи по опытным данным представляет трудную задачу и универсальных способов не существует.

Можно указать простой способ определения вида моделей с одной независимой переменной: строят график полученных экспериментальных данных. Если экспериментальные точки образуют некоторую гладкую кривую, то форма этой кривой может подсказать аналитический вид кривой.

Для проверки согласованности выбранной формы модели с экспериментальными данными преобразуют переменные к линейной форме (модели):

$$y = \alpha + \beta x$$

Пример. Преобразовать переменные к линейной форме.

1. Модель $\eta = \alpha + \beta \dfrac{1}{\gamma}$. Принимаем $x = \dfrac{1}{\gamma}, y = \eta$. Тогда $y = \alpha + \beta x$.

2. $\eta = a\gamma^{\beta}; \log \eta = \log a + \beta \log \gamma$

$y = \alpha + \beta x; \alpha = \log a, x = \log \gamma$

и т.д.

Таким образом, параметры нелинейной модели можно определить по экспериментально статистическому исследованию модифицированной линейной модели. Однако, данный способ трудно применить в тех случаях, когда к выходной переменной модели добавляется ненаблюдаемая ошибка ε .

Например, выражения $\eta = \alpha\gamma^{\beta} + \varepsilon$ и $\log\eta = \log\alpha + \beta\log\gamma + \varepsilon$ неравнозначны, т.к. надо еще прологарифмировать ошибку ε .

При исследовании реального процесса всегда имеется возможность представить результаты экспериментальных данных множеством моделей. Чтобы выбрать одну из них, надо знать для какой *цели* находится модель, и какие *требования* к ней предъявляются. Очевидно, *главным требованием* следует считать способность модели предсказывать направление дальнейших опытов. *Другим требованием* является минимизация суммы квадратов отклонений наблюдаемых значений от вычисленных. Далее, желательно, чтобы преобразованная величина подчинялась нормальному распределению и чтобы дисперсия преобразованной случайной величины не зависела от самой величины.

Важным требованием является простота модели. Если несколько различных моделей отвечают необходимым требованиям, то следует принять ту из них, которая имеет более простую зависимость.

Рассмотрим графическую интерпретацию уравнения модели.

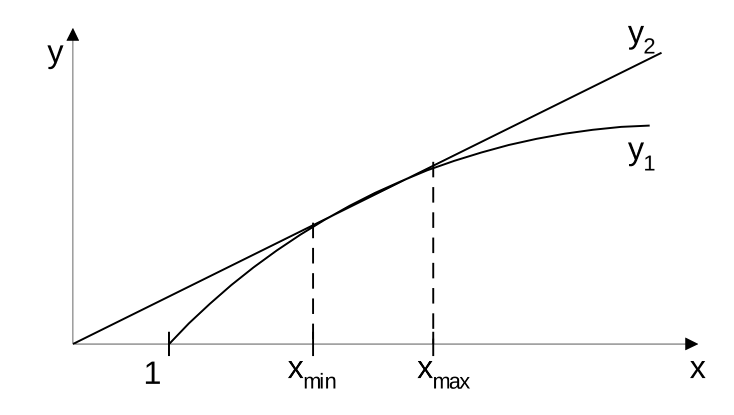


Рисунок 2.2 — Графическая интерпретация уравнения модели

На рисунке 2.2 изображена логарифмическая функция. На отрезке $[x_{min}, x_{max}]$ она с определенной точностью может быть описана двумя уравнениями:

$$y_1 = \log_b x \qquad (2.2)$$

$$y_2 = bx \qquad (2.3)$$

Уравнение (2.3) является линейным алгебраическим и его функциональная зависимость проще, чем уравнение (2.2), которое является трансцендентным.

Поэтому, при прочих равных условиях более предпочтительными являются степенные ряды. Таким образом, выбор модели, как правило, следует искать среди алгебраических полиномов. Причем на первом этапе планирования эксперимента выбирают линейную модель в виде алгебраического полинома первой степени (2.1).

2.3.2 Динамические модели

Динамические модели можно подразделить на два больших класса: *традиционные модели* и *современные модели* или ещё их называют *модели в пространстве состояний* (рис. 2.3).

К ***традиционным моделям*** относят простейшие *инерционные модели*, *модели на базе передаточных функций* и *комплексных коэффициентов передачи.*

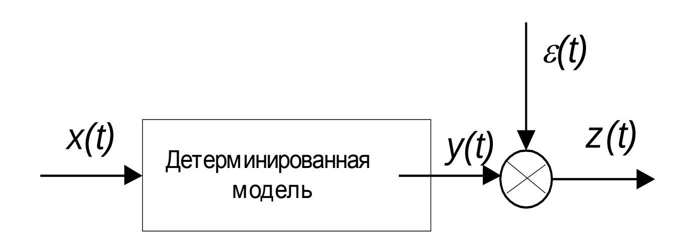


Рисунок 2.3 — Схема динамической модели

Многие технологические процессы обладают инерционностью. Говорят, что такие процессы имеют «память». Инерционные модели бывают дискретные и непрерывные, линейные и нелинейные.

Если *модель линейная дискретная*, то её представляют в виде *суммы свертки* (весовая функция):

$$Y[nT] = \sum_{m=0}^{m-1} g[m\Delta\tau] \cdot x[n\Delta t - (m+1)\Delta\tau] + e[n\Delta t] \qquad (2.4)$$

где $Y[nT]$- дискретные значения выходного сигнала; $g[m\Delta\tau]$- дискретная весовая функция; $x[n\Delta t - (m+1)\Delta\tau]$- дискретные значения входного сигнала; $e[n\Delta t]$- случайная помеха.

В этой модели по существу нет параметров, она непараметрическая. Роль величин, которые необходимо определить из экспериментальных данных, играют значения ординат импульсной характеристики, которые рассматривают как коэффициенты регрессивной модели, а роль факторов здесь играют значения одной и той же входной величины, но в разные моменты времени.

Если *линейная модель непрерывная*, то модель будет типа *интеграла свертки*:

$$y(t) = \int_0^{\infty} g(\tau)x(t-\tau)d\tau + e(t) \qquad (2.5)$$

Модель (2.5) также как и (2.4) является непараметрической. Она содержит неизвестную весовую функцию $g(\tau)$. На практике широко используется возможность представления весовой функции для стационарной системы в *форме Релея-Ритца* путем разложения функции в ряд по системе известных ортогональных функций:

$$g(\tau) = \sum_{i=1}^{k} a_i \varphi_i(\tau,\alpha)$$

где $\varphi_i(\tau,\alpha)$ — заданная система базисных функций (фильтров), зависящая от параметра α.

Такой прием делает модель параметрической. Теперь она содержит ограниченное множество параметров a_i, подлежащих определению.

Нелинейные инерционные модели могут быть представлены в виде сумм или рядов Вольтерра.

В импульсном варианте модель можно представить:

$$Y[nt] = g_0(nT) + \sum_{i=0}^{\infty} g_1(n,i)\cdot x(n-i) + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} g_2(n,i,j)\cdot x(n-i)\cdot x(n-j) \qquad (2.6)$$

Для непрерывного объекта:

$$y(t) = g_0(t) + \int_{\tau_1=0}^{\infty} g_1(t,\tau_1)\cdot x(t-\tau_1)d\tau_1 + \\ + \int_{\tau_1=0}^{\infty}\int_{\tau_2=0}^{\infty} g_2(t,\tau_1,\tau_2)\cdot x(t-\tau_1)\cdot x(t-\tau_2)d\tau_1 d\tau_2 + \ldots + e(t) \qquad (2.7)$$

В выражениях (2.6) и (2.7) g_1 и g_2 соответственно весовые функции, их называют *ядрами Вольтерра первого* и *второго порядка*; g_0 - составляющая не связанная с входным сигналом.

Непосредственное определение ядер по опытным данным представляет собой сложную задачу. Поэтому их обычно аппроксимируют путем разложения в ряд по системе известных ортогональных функций:

$$g_1(t,\tau_1,\alpha) = \sum_{i=1}^{k} a_i \varphi_i(\tau_1,\alpha)$$

$$g_2(t,\tau_1,\tau_2,\alpha) = \sum_i^k \sum_j^k a_{ij} \varphi_i(\tau_1,\alpha) \cdot \varphi_j(\tau_2,\alpha)$$

Теперь задача построения модели сводится к определению параметров весовой функции по косвенным экспериментальным данным.

Модели на базе передаточных функций

Для непрерывных динамических объектов выходная величина определяется интегралом свертки (2.5).

Применим преобразование Лапласа. На основании теоремы свертывания имеем:

$$y(p) = g(p) \cdot x(p) + e(p) \tag{2.8}$$

Для импульсного динамического объекта, согласно выражения (4.4) после z- преобразования получаем:

$$y(z) = g(z) \cdot x(z) + e(z) \tag{2.9}$$

Модели (2.8) и (2.9) имеют вид регрессионных зависимостей. Параметры этих моделей, которые необходимо определить из экспериментальных данных, содержатся в выражениях передаточных функций.

Модели на основе комплексного коэффициента передачи

К модели частотного вида можно перейти путем замены в выражении (2.8) аргумента p на jw:

$$y(jw) = g(jw) \cdot x(jw) + e(jw) \tag{2.10}$$

В роли аргументов в моделях (2.8), (2.9), (2.10) выступает не время, а соответственно параметры преобразования: *z, p, jw*. Все эти модели линейны по входным сигналам, но, как правило, не линейны по параметрам.

Современные модели (модели в пространстве состояний)

В настоящее время все чаще переходят к математическому описанию динамических систем в описание систем в пространстве состояний. При этом используют *конечно — разностные* и *дифференциальные уравнения* в форме Коши, т.е. разрешенных относительно первых разностей и первых производных. Описание систем в пространстве состояний позволяет с единых позиций рассматривать различные системы: линейные, нелинейные, дискретные и непрерывные.

Модели в виде конечно —разностных уравнений

Пусть динамический объект описывается линейным конечно-разностным уравнением *l*-го порядка с постоянными коэффициентами. Далее считаем, что наблюдения производятся в дискретные равноотстоящие моменты времени. Тогда модель можно представить выражением (2.4).

Примером таких моделей являются дискретно — непрерывные

модели.

Модели в виде обыкновенных дифференциальных уравнений

Обыкновенное линейное дифференциальное уравнение *l*-го порядка с постоянными коэффициентами имеет вид:

$$c_0 \frac{d^l y}{dt^l} + c_1 \frac{d^{l-1} y}{dt^{l-1}} + c_2 \frac{d^{l-2} y}{dt^{l-2}} + \ldots + c_l y = d_0 \frac{d^q x}{dt^q} + d_1 \frac{d^{q-1} x}{dt^{q-1}} + \ldots + d_q x \qquad (2.11)$$

Уравнением (2.11) описываются рассмотренные выше непрерывно- детерминированные модели (D- схемы).

Контрольные вопросы

1. Какие способы представления моделирующего алгоритма Вы знаете?
2. В чем заключаются принципы фиксированного и переменного шага?
3. В чем заключается принцип особых состояний?
4. В чем заключается принцип последовательной проводки заявок?
5. Перечислите отличия специализированных языков от универсальных языков.
6. Какие виды моделей в статике Вы знаете? Перечислите их особенности.
7. Какие виды динамических моделей Вы знаете? Перечислите их особенности.

3. План эксперимента

Будем считать, что математическая модель выбрана и представляет собой зависимость

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n$$

Выходная величина η может иметь и более сложную форму переменных x_i.

Если поставлена задача найти оптимальные условия для η, то выходную величину называют ***параметром оптимизации***.

В качестве выходного параметра η может быть температура, давление или более сложный показатель качества выпускаемой продукции.

В тоже время в планировании могут участвовать сложные факторы, например: соотношения между компонентами; соотношения, определяемые алгоритмом функционирования и т.д.

С помощью сложных факторов можно учесть некоторые динамические свойства объекта, исследуя его в статическом режиме. Например, если какое-то измерение во времени имеет сложный вид кривой, то целесообразно в качестве фактора обозначить номер кривой. Тогда различные варианты кривых будут рассматриваться как уровни. Таким образом можем представить сложный фактор — функцию с помощью простых однозначных факторов.

Выбор экспериментальной области факторного пространства связан с тщательным анализом априорной информации. В этой области находят локальную подобласть для планирования эксперимента.

При наличии априорных сведений о процессе за основной уровень принимаются те значения факторов, при которых выходная величина принимает лучшее значение. Построение плана эксперимента сводится к выбору экспериментальных точек, симметричных относительно основного уровня.

3.1 Выбор интервалов варьирования

После выбора основного уровня выбирают интервал варьирования для каждого из факторов.

Интервалом варьирования факторов называют некоторое число, прибавление которого к основному уровню дает верхний, а вычитание — нижний уровни факторов. То есть это расстояние на координатной оси между основным и верхним (либо нижним) уровнем.

Для упрощения записи условий эксперимента и обработки экспериментальных данных масштабы по осям выбираются таким образом, чтобы верхний уровень соответствовал +1, а нижний -1, основной — нулю. Уровни, записанные в таких обозначениях,

называются ***кодированными*** :

$$x_j = \frac{x_{j'} - x_{jo}}{\lambda_j} \qquad (3.1)$$

где x_j — кодированное значение фактора; j — номер фактора; $x_{j'}$— натуральное значение фактора; x_{jo} — натуральное значение основного уровня; λ_j — интервал варьирования.

Пример. Дано интервал варьирования λ_1=2; натуральное значение основного уровня x_{1o}=3 (рис. 3.1)

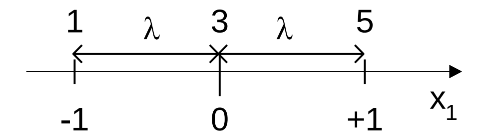


Рисунок 3.1 — Кодирование уровней

Верхний кодированный уровень: $x_в = \frac{5-3}{2} = 1$; нижний кодированный уровень: $x_н = \frac{1-3}{2} = -1$.

Экспериментальные планы, в которых все факторы выбираются на двух уровнях, называются ***планами типа 2^K^***, где K — число факторов (рис. 3.1). Встречаются планы 3^K и более числа уровней.

На выбор интервала варьирования накладываются естественные ограничения сверху и снизу. Интервал варьирования не может быть меньше ошибки экспериментатора, которую он допускает при фиксации уровня фактора. В противном случае верхний и нижний уровни будут неразличимы. С другой стороны интервал варьирования ограничен сверху пределами области определения.

После определения интервалов варьирования выбирают значения факторов, т.е. число уровней для каждого фактора. Число уровней всегда ограничено снизу и не может быть меньше двух, т.е. если принять один уровень, то фактор окажется постоянным во всех опытах.

Рандомизацией называется любая процедура, обеспечивающая случайный порядок проведения опытов. Такую процедуру осуществляют с помощью случайного выбора номера опыта. При проведении эксперимента всегда следует применять рандомизацию.

3.2 Полный факторный эксперимент

Эксперимент, в котором реализуются все возможные сочетания

уровней факторов, называют ***полным факторным экспериментом*** (ПФЭ). При двух уровнях имеем ПФЭ типа 2^K. Число опытов для данного случая будет равно

$$N = 2^K$$

Условие эксперимента записываются в виде таблицы. Строки её соответствуют различным опытам (вектор — строка), столбцы — значениям факторов в кодированном виде (вектор — столбцы). Такие таблицы называются ***матрицами планирования эксперимента*** (МПЭ).

Составим МПЭ для двумерной модели

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \qquad (3.2)$$

на двух уровнях 2^2 (табл.3.1). Число опытов $N=2^2=4$.

Таблица 3.1

МПЭ 2^2

Опыт	x_1	x_2	y
1	*-1*	*-1*	y_1
2	*+1*	*-1*	y_2
3	*-1*	*+1*	y_3
4	*+1*	*+1*	y_4

План эксперимента можно представить геометрически (рис. 3.2). Для плана 2^2 каждая комбинация факторов представляет собой вершину квадрата.

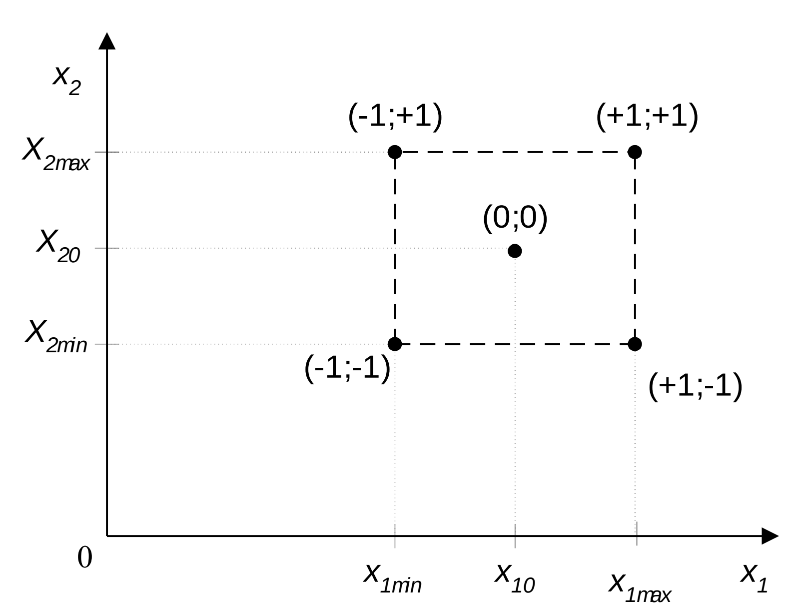


Рисунок 3.2 — Геометрическое представление ПФЭ

В области определения факторов находят точку соответствующую основному уровню. Через эту точку проводят новые оси координат, параллельно осям натуральных значений

факторов. Затем выбирают масштабы по новым осям для каждого фактора согласно выражению (3.1).

В МПЭ вводится фиктивный столбец x_0 для учета свободного члена β_0.

Коэффициенты $\beta_0, \beta_1, \beta_2$ оцениваются согласно выражений:

$$b_0 = \frac{\sum y_i}{N}, b_1 = \frac{\sum x_{1i} y_i}{N}, b_2 = \frac{\sum x_{2i} y_i}{N}.$$

3.2.1 Свойства полного факторного эксперимента $2^К$

К свойствам МПЭ относятся те, которые определяют качество модели, т.е. эти свойства делают оценки коэффициентов модели наилучшими. Первые два свойства вытекают из построения матрицы.

1. Симметричность относительно центра эксперимента. Алгебраическая сумма элементов столбца каждого фактора равно нулю $\sum_{i=1}^{N} x_{ji} = 0$, где *j*- номер фактора, *N* - число опытов.

2. Условие нормировки. Сумма квадратов элементов каждого столбца равна числу опытов $\sum_{i=1}^{N} x_{ji}^2 = N$.

3. Ортогональность матрицы. Сумма почленных произведений любых двух векторов — столбцов матрицы равна нулю $\sum_{i=1}^{N} x_{ji} x_{ui} = 0$, где $j \neq u; j, u = 0,1,\ldots,k$.

Ортогональные планы делают эксперимент более эффективным.

Ортогональность плана позволяет получить оценки для коэффициентов уравнения регрессии независимые друг от друга. Иными словами ортогональность характеризует отсутствие корреляции между факторами. Однако если имеет место нелинейность, то столбцы взаимодействий окажутся неразличимы, закоррелироваными с некоторыми столбцами линейных эффектов. Это приводит к тому, что по результатам данного эксперимента становится невозможным разделить коэффициенты регрессии между линейными и нелинейными факторами.

4. Рототабельность планов. Это такие планы, для которых дисперсия $\hat{y}$ одинакова для всех точек пространства переменных *x*, лежащих на одинаковых расстояниях от центра.

3.2.2 Выбор модели при проведении полного факторного эксперимента

Планируя эксперимент на первом этапе, всегда стремятся получить линейную модель. Для двух факторов модель представляют в виде выражения (3.2). Однако не всегда экспериментатор имеет

гарантии, что в выбранных интервалах варьирования процесс описывается линейной моделью. Часто встречающийся вид нелинейности связан с эффектом взаимодействия между факторами. ПФЭ позволяет оценить кроме коэффициентов при линейных эффектах коэффициенты взаимодействия. Для этого перемножают соответствующие столбцы. Тогда уравнение принимает вид

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \tag{3.3}$$

МПЭ с учетом *фактора взаимодействия* для ПФЭ 2^2 показана в таблице 3.2.

Таблица 3.2

МПЭ для ПФЭ 2^2

Опыт	x_0	x_1	x_2	x_1x_2	y
1	*+1*	*-1*	*-1*	*+1*	y_1
2	*+1*	*+1*	*-1*	*-1*	y_2
3	*+1*	*-1*	*+1*	*-1*	y_3
4	*+1*	*+1*	*+1*	*+1*	y_4

Коэффициенты уравнений регрессии (3.2, 3.3) оцениваются следующим образом:

$$\beta_0 \to b_0 = \frac{\sum_{i=1}^{N} y_i}{N}, \beta_j \to b_j = \frac{\sum x_{ji} y_i}{N}, \beta_{12} \to b_{12} = \frac{\sum x_{ji} x_{ui} y_i}{N}, j \neq u$$

По столбцам x_1 и x_2 осуществляют планирование, что же касается столбцов , x_0 и x_1x_2 ,то они служат только для расчета.

При составлении МПЭ руководствуются следующими правилами:

- располагают, если имеется соответствующая информация, факторы в матрице в порядке убывания степени их влияния на целую функцию;
- стремятся выполнить требования рандомизации варьирования уровней.

При составлении матрицы уменьшают частоту чередования уровней при переходе от x_1 к x_2, от x_2 к x_3 и т.д. каждый раз вдвое.

Рассмотрим пример составления МПЭ для трех факторного полного эксперимента. В качестве уравнения регрессии берем неполную квадратичную модель.

$$\hat{y} = b_0 + \sum_{i=1}^{3} b_i x_i + \sum_{i<j}^{3} b_{ij} x_i x_j + b_{123} x_1 x_2 x_3 \tag{3.4}$$

Введем обозначение переменных x через z, тогда

$$\hat{y} = \sum_{i=1}^{z} b_i z_i \tag{3.5}$$

где $\sum_{i=1}^{3} b_i x_i = \sum_{i=1}^{3} b_i z_i$, $\sum_{i<j}^{3} b_{ij} x_i x_j = \sum_{i=4}^{6} b_i z_i$, $b_{123} x_1 x_2 x_3 = b_7 z_7$.

Составим МПЭ. $N = 2^3 = 8$ (табл. 3.3).

Таблица 3.3

МПЭ для ПФЭ 2^3

Номер опыта	x_0 / z_0	x_1 / z_1	x_2 / z_2	x_3 / z_3	x_1x_2 / z_4	x_1x_3 / z_5	x_2x_3 / z_6	$x_1x_2x_3$ / z_7	*Код. обозначение*
1	+1	-1	-1	-1	+1	+1	+1	-1	y_1
2	+1	+1	-1	-1	-1	-1	+1	+1	y_2
3	+1	-1	+1	-1	-1	+1	-1	+1	y_3
4	+1	+1	+1	-1	+1	-1	-1	-1	y_4
5	+1	-1	-1	+1	+1	-1	-1	+1	y_5
6	+1	+1	-1	+1	-1	+1	-1	-1	y_6
7	+1	-1	+1	+1	-1	-1	+1	-1	y_7
8	+1	+1	+1	+1	+1	+	+1	+1	y_8

После экспериментирования проверяют однородность дисперсии, значимость коэффициентов и адекватность модели.

3.3 Дробный факторный эксперимент

Во многих реальных процессах некоторые факторы взаимодействия могут отсутствовать. И тогда ПФЭ будет обладать избыточностью опытов.

Рассмотрим пути минимизации числа опытов.

Обратимся к уравнению (3.3). Если располагаем сведениями о том, что в выбранных интервалах варьирования процесс в статике может быть описан линейной моделью, то достаточно определить три коэффициента b_0 ,b_1 ,b_2. В результате остается одна степень свободы, т.к. имеем четыре опыта, а количество констант три. Используем эту степень свободы для минимизации числа опытов. При линейном приближении $b_{12} \to 0$ и тогда вектор — столбец x_1x_2 может быть использован для нового фактора x_3.

При этом эксперименте появляются смешанные оценки

$$b_1 \to \beta_1 + \beta_{23}, b_2 \to \beta_2 + \beta_{13}, b_3 \to \beta_3 + \beta_{12}, \qquad (3.6)$$

т.е. столбцы.

Таблица 3.4

МПЭ для ДФЭ 2^{3-1}

Опыт	x_0	x_1	x_2	x_3	y
1	+1	+1	+1	+1	y_1
2	+1	-1	+1	-1	y_2
3	+1	+1	-1	-1	y_3
4	+1	-1	-1	+1	y_4

Пример. Допустим x_1 и x_2x_3 между собой неразличимы. Однако парные взаимодействия в линейной модели незначительны. Зато вместо восьми опытов для изучения влияния трех факторов можно поставить только четыре опыта, т.е. вместо ПФЭ 2^3 получаем 2^{3-1}. В теории эксперимента 2^{3-1} называют ***полу —репликой***. В общем случае имеют дело с дробной репликой. А факторный эксперимент называют ***дробным*** (ДФЭ).

Для уяснения принципа составления МПЭ ДФЭ введено понятие *определяющего контраста*. Он позволяет определить, какие оценки смешаны друг с другом, не изучая МПЭ для выявления совпадающих столбцов. Для этого берут символичное обозначение произведения столбцов равного +1 или -1. Это и называют *контрастом*. Чтобы определить какой эффект смешан с данным, нужно помножить обе части определяющего контраста на столбец, соответствующий данному эффекту.

Пример. Пусть имеем три фактора x_1 ,x_2 ,x_3 . При построении полуреплики 2^{3-1} имеется только две возможности приравнять x_3 либо к «$+x_1x_2$», либо к «$-x_1x_2$» (табл. 3.5).

Таблица 3.5

МПЭ для ДФЭ 2^{3-1}

Опыт	x_1	x_2	x_3	$x_1x_2x_3$	*Опыт*	x_1	x_2	x_3	$x_1x_2x_3$
1	*-1*	*-1*	*+1*	*+1*	*1*	*-1*	*-1*	*-1*	*-1*
2	*+1*	*-1*	*-1*	*+1*	*2*	*+1*	*-1*	*+1*	*-1*
3	*-1*	*+1*	*-1*	*+1*	*3*	*-1*	*+1*	*+1*	*-1*
4	*+1*	*+1*	*+1*	*+1*	*4*	*+1*	*+1*	*-1*	*-1*

Возьмем в качестве определяющего контраста $-1 = x_1x_2x_3$.Тогда $-x_1 = x_1^2x_2x_3$. Учитывая, что $x^2 = 1$ получаем $x_1 = -x_2x_3$.

Теперь возьмем за определяющий контраст $+1 = x_1x_2x_3$. Получаем: $x_1 = x_2x_3, x_2 = x_1x_3, x_3 = x_1x_2$. Эти выражения показывают, что коэффициенты линейного уравнения будут оценками (3.6).

Соотношение, показывающее с какими из эффектов смешан данный эффект, называется ***генерирующим соотношением***.

При выборе полуреплики 2^{4-1} возможны восемь генерирующих соотношений:

1. $x_4 = x_1x_2$
2. $x_4 = -x_1x_2$
3. $x_4 = x_2x_3$
4. $x_4 = -x_2x_3$
5. $x_4 = x_1x_3$
6. $x_4 = -x_1x_3$
7. $x_4 = x_1x_2x_3$
8. $x_4 = -x_1x_2x_3$

Разрешающая способность этих полуреплик различна. Реплики

1 — 6 имеют по три фактора и носят название ***планов с расширяющей способностью III*** (по наибольшему числу факторов в определяющем контрасте). Реплики 7 — 8 имеют по четыре фактора и обладают максимальной разрешающей способностью. Их называют ***главными репликами***. Всегда стремятся выбрать реплику с наибольшей разрешающей способностью, т.к. чем больше эффектов взаимосвязано, тем точнее окажется полученная модель.

Однако, если имеется информация об эффектах взаимодействия, то реплики нужно выбирать с ее учетом.

Реализация МПЭ ДФЭ ничем не отличается от реализации МПЭ ПФЭ. Методика оценки значимости коэффициентов и проверка адекватности модели проводится также как и в ПФЭ.

3.4 Обобщающий определяющий контраст

Рассмотрим на примере исследование модели с пятью факторами. Возьмём реплику 2^{5-2}. Получаем 8 опытов вместо 32.

Возможны 12 решений, если приравнять x_4 парному взаимодействию, а x_5 — тройному:

1.	$x_4=x_1x_2$	$x_5=x_1x_2x_3$
2.	$x_4=x_1x_2$	$x_5=-x_1x_2x_3$
3.	$x_4=-x_1x_2$	$x_5=x_1x_2x_3$
4.	$x_4=-x_1x_2$	$x_5=-x_1x_2x_3$
5.	$x_4=x_1x_3$	$x_5=x_1x_2x_3$
6.	$x_4=x_1x_3$	$x_5=-x_1x_2x_3$
7.	$x_4=-x_1x_3$	$x_5=x_1x_2x_3$
8.	$x_4=-x_1x_3$	$x_5=-x_1x_2x_3$
9.	$x_4=x_2x_3$	$x_5=x_1x_2x_3$
10.	$x_4=x_2x_3$	$x_5=-x_1x_2x_3$
11.	$x_4=-x_2x_3$	$x_5=x_1x_2x_3$
12.	$x_4=-x_2x_3$	$x_5=-x_1x_2x_3$

Допустим выбран первый вариант. Тогда определяющими контрастами будут: $1=x_1x_2x_4$, $1=x_1x_2x_3x_5$. Перемножим эти определяющие константы, получим третье соотношение: $1=x_3x_4x_5$. Для того чтобы полностью охарактеризовать разрешающую способность реплики, вводят понятие ***обобщающего определяющего контраста***: $1= x_1x_2x_4= x_3x_4x_5=x_1x_2x_3x_5$.

Система смешивания столбца определяется умножением обобщающего определяющего контраста последовательно на x_1,x_2,x_3:

$$x_1=x_2x_4=x_1x_3x_4x_5=x_2x_3x_5;$$

$$x_2=x_1x_4=x_2x_3x_4x_5=x_1x_3x_5;$$

$$x_3 = x_1x_2x_3x_4= x_4x_5x_1x_2x_5;$$

$$x_4=x_1x_2=x_3x_5=x_1x_2x_3x_4x_5;$$

$$x_5=x_1x_2x_4x_5=x_3x_4=x_1x_2x_3;$$

$$x_1x_2=x_4=x_1x_2x_3x_4x_5=x_3x_5.$$

Если при выбранной реплике некоторые коэффициенты получаются отличными от нуля, например:

$$b_{12} \to \beta_{12} + \beta_4 + \beta_{35} + \beta_{12345}$$

то ставят вторую серию опытов с другой репликой, например, берут вариант 4.

Дробные реплики находят широкое применение при получении линейных моделей, причем, целесообразность применения их возрастает с ростом количества факторов. Эффективность применения дробных реплик зависит от выбора системы смешивания линейных эффектов с эффектами взаимодействия.

3.5 Планирование экспериментов при построении полной квадратичной модели

В уравнениях (3.2), (3.3), (3.4) учитывались только линейные эффекты и эффекты взаимодействия. В некоторых случаях существенными могут оказаться коэффициенты при квадратных переменных, их кубов и т.д.

Для двухфакторного эксперимента модель может быть представлена выражением

$$y = b_0x_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2 \qquad (3.7)$$

Полученные вектор - столбцы x_1^2 и x_2^2 являются единичными столбцами, совпадающие друг с другом и с фиктивным столбцом x_0. Эти столбцы неразличимы, поэтому нельзя сказать за счет чего получилась величина b_0. Очевидно, она включает в себя значения свободного члена β_0 и вклады квадратичных членов. Символически это можно записать:

$$b_0 \to \beta_0 + \sum_{i=1}^{k}\beta_{ii}$$

Для квадратичной модели получается следующая система смешивания:

$$b_0 \to \beta_0 + \beta_{11} + \beta_{22}, b_1 \to \beta_1, b_2 \to \beta_2, b_{12} \to \beta_{12}.$$

Следовательно, планирование эксперимента на двух уровнях не дает возможности получить раздельные оценки коэффициентов при квадратичных членах и фиктивной переменной x_0.

Согласно теории интерполяции, для решения задачи нахождения раздельных оценок число уровней каждой из независимых переменных должно быть на единицу больше степени интерполяционного полинома. Для полинома второй степени число уровней должно быть равно трем.

Однако применение методов ПФЭ плана 3^n не является

рациональным из-за резкого увеличения опытов эксперимента. Поэтому разработаны специальные методы построения планов второго порядка.

Например, в качестве двухфакторных планов второго порядка могут служить планы, представляемые вершинами и, по крайней мере, одной центральной точкой любого *(n-1)* мерного правильного многоугольника (который можно вписать в круг).

Пример. Имеем восьмиугольный план (рис. 3.3, табл.3.6).

Этот пример можно обобщить на случай получения планов второго порядка. Для этого к ПФЭ типа 2^n добавляется центральная точка с координатами (0,0,...0) и, так называемые, *звёздные точки* с координатами (0,0,..., $\pm\alpha$,...,0), лежащие на сфере диаметра 2α. Т.е. план ПФЭ достраивается до *плана второго порядка*. Такой план называется ***композиционным планом.***

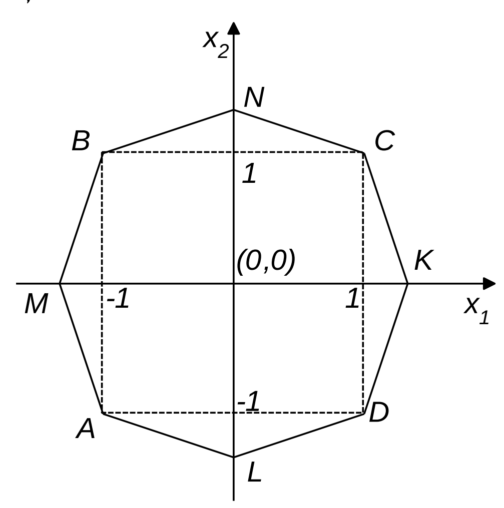


Рисунок 3.3 — Восьмиугольный план эксперимента

Таблица 3.6

Матрица композиционного плана двухфакторной модели

Опыт	x_1	x_2	*Описание*
1	*-1*	*-1*	*План 2^2 представлен квадратом АВСД*
2	*+1*	*-1*	
3	*-1*	*+1*	
4	*+1*	*+1*	
5	$\sqrt{2}$	*0*	*План представлен звёздными точками MNKL*
6	$-\sqrt{2}$	*0*	
7	*0*	$\sqrt{2}$	
8	*0*	$-\sqrt{2}$	
9	*0*	*0*	*Центральная точка*

Для трехфакторной модели МПЭ будет иметь вид, представленный в таблице 3.7.

Таблица 3.7

Матрица композиционного плана двухфакторной модели

Опыт	x_0	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	*Описание*
1	+1	-1	-1	-1	+1	+1	+1	
2	+1	+1	-1	-1	+1	+1	+1	*ПФЭ*
3	+1	-1	+1	-1	+1	+1	+1	
4	+1	+1	+1	-1	+1	+1	+1	2^2
5	+1	-1	-1	+1	+1	+1	+1	
6	+1	+1	-1	+1	+1	+1	+1	
7	+1	-1	+1	+1	+1	+1	+1	
8	+1	+1	+1	+1	+1	+1	+1	
9	+1	$-\alpha$	0	0	α^2	0	0	
10	+1	$+\alpha$	0	0	α^2	0	0	*Звёздные*
11	+1	0	$-\alpha$	0	0	α^2	0	
12	+1	0	$+\alpha$	0	0	α^2	0	*точки*
13	+1	0	0	$-\alpha$	0	0	α^2	
14	+1	0	0	$+\alpha$	0	0	α^2	
15	+1	0	0	0	0	0	0	
16	+1	0	0	0	0	0	0	*Центр. точка*

Добавление двух сфер, образованных звездными точками и центральной точкой, к ПФЭ позволяет получить раздельные оценки b_0 и b_{ii}. Все три сферы образуют композиционный план второго порядка.

В зависимости от критерия оптимальности плана, различают *ортогональное композиционное планирование и рототабельное композиционное планирование.*

План, приведенный в таблице 3.6, является рототабельным и обеспечивает получение раздельных оценок b_0 и b_{ii}.

3.5.1 Ортогональное центральное композиционное планирование

Критерием оптимальности является ортогональность столбцов матрицы планирования. В силу этого свойства все коэффициенты модели определяются независимо друг от друга.

Ортогональность столбцов x_0 и x_i^2 достигается путем преобразования вида:

$$\hat{x}_i^2 = x_i^2 - \frac{\sum_{j=1}^{N} x_{ji}^2}{N} = x_i^2 - \bar{x}_i^2 \qquad (3.8)$$

С учетом выражения (3.8) условие ортогональности выполняется:

$$\sum_{j=1}^{N} x_{j0}^2 \cdot \hat{x}_{ji}^2 = \sum_{j=1}^{N} x_{j0}^2 [x_{ji}^2 - \frac{\sum_{j=1}^{N} x_{ji}^2}{N}] = \sum_{j=1}^{N} 1 \cdot x_{ji}^2 - \frac{N\sum_{j=1}^{N} x_{ji}^2}{N} =$$

$$= \sum_{j=1}^{N} x_{ji}^2 - N\frac{\sum_{j=1}^{N} x_{ji}^2}{N} = 0$$

Из условия ортогональности выбирают координату α звездной точки для n — независимых переменных.

Таблица 3.8 соответствует ортогональному композиционному плану для количества переменных n=2,3,4 , где N_α — число звездных точек; N_0 — число точек в центре эксперимента; N_c — количество точек куба (гиперкуба) при ПФЭ; N — общее число точек факторного пространства.

Таблица 3.8

Количество точек ортогонального композиционного плана

n	α	N_α	N_0	N_c	N
2	1,0	4	1	4	9
3	1,215	6	1	8	15
4	1,414	8	1	16	25

Составим матрицу ортогонального планирования для трехфакторного эксперимента.

В таблице 3.9 для точек ПФЭ

$$x_i^2 = 1;$$

$$\hat{x}_i^2 = x_i^2 - \bar{x}_i^2 = x_i^2 - 0{,}73.$$

Таблица 3.9

Матрица ортогонального планирования для трехфакторного эксперимента

Номер опыта	x_0	x_1	x_2	x_3	$\hat{x}_1^2$	$\hat{x}_2^2$	$\hat{x}_3^2$	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
	z_0	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
1	+1	-1	-1	-1	0,27	0,27	0,27	+1	+1	+1	-1
2	+1	+1	-1	-1	0,27	0,27	0,27	-1	-1	+1	+1
3	+1	-1	+1	-1	0,27	0,27	0,27	-1	+1	-1	+1
4	+1	+1	+1	-1	0,27	0,27	0,27	+1	-1	-1	-1
5	+1	-1	-1	+1	0,27	0,27	0,27	+1	-1	-1	+1
6	+1	+1	-1	+1	0,27	0,27	0,27	-1	+1	-1	-1

7	*+1*	*-1*	*+1*	*+1*	*0,27*	*0,27*	*0,27*	*-1*	*-1*	*+1*	*-1*
8	*+1*	*+1*	*+1*	*+1*	*0,27*	*0,27*	*0,27*	*+1*	*+1*	*+1*	*+1*
9	*+1*	*-1,215*	*0*	*0*	*0,75*	*-0,73*	*-0,73*	*0*	*0*	*0*	*0*
10	*+1*	*+1,215*	*0*	*0*	*0,75*	*-0,73*	*-0,73*	*0*	*0*	*0*	*0*
11	*+1*	*0*	*-1,215*	*0*	*-0,73*	*0,75*	*-0,73*	*0*	*0*	*0*	*0*
12	*+1*	*0*	*+1,215*	*0*	*-0,73*	*0,75*	*-0,73*	*0*	*0*	*0*	*0*
13	*+1*	*0*	*0*	*-1,215*	*-0,73*	*-0,73*	*0,75*	*0*	*0*	*0*	*0*
14	*+1*	*0*	*0*	*+1,215*	*-0,73*	*-0,73*	*0,75*	*0*	*0*	*0*	*0*
15	*+1*	*0*	*0*	*0*	*-0,73*	*-0,73*	*-0,73*	*0*	*0*	*0*	*0*

Где 1-8 опыт соответствует ПФЭ 2^3; 9-14 опыт соответствует звездным точкам; 15 опыт - центральная точка.

Уравнение регрессии для полного квадратичного полинома будет иметь вид:

$$y = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i,j=1}^{n} b_{ij} x_i x_j + \sum_{i=1}^{n} b_{ii} (x_i^2 - \overline{x}_i^2) =$$

$$= b_0^* + \sum_{i=1}^{n} b_i x_i + \sum_{i,i=1}^{n} b_{ij} x_i x_j + \sum_{i=1}^{n} b_{ii} x_i^2, \qquad (3.9)$$

где $b_0^* = b_0 - \sum_{i=1}^{n} b_{ii} \overline{x}_i^2$, n- количество факторов.

Дисперсия коэффициента b_0^* будет оцениваться по формуле:

$$S_{b_0^*}^2 = S_{b_0}^2 + \sum_{i=1}^{n} S_{b_{ii}}^2 \overline{x}_i^2 \qquad (3.10)$$

Коэффициенты b_i, b_{ii} и b_{ij} определяются по формулам:

$$b_i = \frac{\sum_{j=1}^{N} x_{ji} \overline{y}_j}{\sum_{i=1}^{n} x_{ji}}, \quad b_{ii} = \frac{\sum_{u=1}^{N} x^2_{ui} \overline{y}_u}{N}, \quad b_{ij} = \frac{\sum_{u=1}^{N} x_{ui} x_{uj} \overline{y}_u}{N},$$

где $u = \overline{1, N}$ - число опытов; $\overline{y}_u = \frac{\sum_{p=1}^{P} y_p}{P}$, P- число прогонов.

Из-за неодинаковой дисперсии коэффициентов регрессии критерий ортогональности является недостаточно сильным критерием оптимальности для планирования второго порядка. Поэтому точность предсказания выходной величины в различных направлениях факторного пространства неодинакова.

Лучшим методом планирования является такой метод, который обеспечивает одинаковую точность во всех направлениях на одинаковом расстоянии от центра. Таким методом является *рототабельное композиционное планирование*.

3.5.2 Ротатабельное композиционное планирование

Критерием оптимальности в ротатабельном планировании является условие $\sigma_y^2 = const$ при одинаковом удалении точек эксперимента от центра, т.е. $R = const$.

Если имеются двухфакторные планы, то, как уже было отмечено, типичными примерами ротатабельных планов являются планы, представляемые вершинами и, по крайней мере, одной центральной точкой любого *(n-1)* — мерного правильного многоугольника, который можно вписать в круг (рис. 3.3).

Композиционные центральные ротатабельные планы также как и ортогональные состоят из трех сфер: сфера нулевого радиуса - центральные точки; сфера точек куба или гиперкуба и сфера звездных точек. Равномерность расположения точек на сфере приводит к вырожденным матрицам. Для устранения вырожденности используют сферу нулевого радиуса с несколькими центральными точками.

Таблица 3.10

Количество точек ротатабельного композиционного плана

n	α	N_α	N_0	N_c	N
2	*1,414*	*4*	*5*	*4*	*13*
3	*1,682*	*6*	*6*	*8*	*20*
4	*2*	*8*	*7*	*16*	*31*

где N_α - число звездных точек; N_0 - число точек в центре эксперимента; N_c - количество точек куба (гиперкуба); N - общее число точек факторного пространства.

Матрица планирования ротатабельного плана второго порядка для трехфакторного эксперимента будет представлена в таблице 3.11.

Таблица 3.11

Матрица ротатабельного планирования для трехфакторного эксперимента

Номер	x_0	x_1	x_2	x_3	x_1^2	x_2^2	x_3^2	x_1x_2	x_1x_3	x_2x_3
опыта	z_0	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9
1	*+1*	*-1*	*-1*	*-1*	*+1*	*+1*	*+1*	*+1*	*+1*	*+1*
2	*+1*	*+1*	*-1*	*-1*	*+1*	*+1*	*+1*	*-1*	*-1*	*+1*
3	*+1*	*-1*	*+1*	*-1*	*+1*	*+1*	*+1*	*-1*	*+1*	*-1*
4	*+1*	*+1*	*+1*	*-1*	*+1*	*+1*	*+1*	*+1*	*-1*	*-1*
5	*+1*	*-1*	*-1*	*+1*	*+1*	*+1*	*+1*	*+1*	*-1*	*-1*
6	*+1*	*+1*	*-1*	*+1*	*+1*	*+1*	*+1*	*-1*	*+1*	*-1*
7	*+1*	*-1*	*+1*	*+1*	*+1*	*+1*	*+1*	*-1*	*-1*	*+1*
8	*+1*	*+1*	*+1*	*+1*	*+1*	*+1*	*+1*	*+1*	*+1*	*+1*

9	*+1*	*-1,682*	*0*	*0*	*2,828*	*0*	*0*	*0*	*0*	*0*
10	*+1*	*+1,682*	*0*	*0*	*2,828*	*0*	*0*	*0*	*0*	*0*
11	*+1*	*0*	*-1,682*	*0*	*0*	*2,828*	*0*	*0*	*0*	*0*
12	*+1*	*0*	*+1,682*	*0*	*0*	*2,828*	*0*	*0*	*0*	*0*
13	*+1*	*0*	*0*	*-1,682*	*0*	*0*	*2,828*	*0*	*0*	*0*
14	*+1*	*0*	*0*	*+1,682*	*0*	*0*	*2,828*	*0*	*0*	*0*
15	*+1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
16	*+1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
17	*+1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
18	*+1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
19	*+1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
20	*+1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*

Эксперимент проводится аналогично ПФЭ, однако оценки коэффициентов рассчитываются по своим формулам:

$$b_0 = \frac{A}{N}[2\lambda^2(n+2)\sum_{j=1}^{N} x_{j,0}\bar{y}_j - 2\lambda C\sum_{i=1}^{n}\sum_{j=1}^{N} x_{j,i}\bar{x}_j] \qquad (3.11)$$

$$b_{ii} = \frac{A}{N}\left\{C^2[(n+2)\lambda - n]\sum_{j=1}^{N} x_{j,i}\bar{y}_j + C^2(1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{N} x_{j,i}\bar{x}_j - 2\lambda C\sum_{j=1}^{N} x_{j,0}\bar{y}_j\right\}$$

(3.12)

$$b_i = \frac{C}{N}\sum_{j=1}^{N} x_{j,i}\bar{y}_j \qquad (3.13)$$

$$b_{ij} = \frac{C^2}{N\lambda}\sum_{u=1}^{N} x_{ui}x_{uj}\bar{y}_u \qquad (3.14)$$

$$C = \frac{N}{\sum\limits_{j=1}^{N} x_{j,i}},\ A = \frac{1}{2\lambda[(n+2)(\lambda-n)]},\ \lambda = \frac{nN\sum\limits_{w=1}^{k} N_w P_w}{(n+2)(\sum\limits_{w=1}^{k} N_w P_w^2)^2} \qquad (3.15 \div 3.17)$$

где N_w - число точек на сфере радиуса P_w; *к* - число сфер (*к=3*).

Рототабельные планы нашли широкое применение на практике. Однако с точки зрения математиков, занимающихся развитием математической статистики, выбор такого критерия представляется *мало обоснованным*. Он не вытекает логически из тех идей, на которые базируется математика. Например, как выбрать расстояние α до звездных точек; не все пространства независимых переменных, отведенное для эксперимента, используется в композиционных планах и т.д. Невозможно из множества рототабельных планов при одном и том же числе факторов выбрать лучший, т.к. не было критерия оценки. Поэтому этот критерий стали относить к *эмпирико-интуитивным критериям*.

Наряду с развитием планирования экспериментов, основанных на эмпирико-интуитивным критериях Бокса в США стало развиваться второе чисто теоретическое направление, которое связывают с планом ученого Кифера. Он установил связь между некоторыми критериями оптимальности; теоретически доказал, что для отдельных видов регрессии одни и те же планы могут отвечать сразу нескольким критериям оптимальности.

3.6 Разбиение матрицы планирования $2^К$ на блоки

При проведении серии экспериментов, осуществляемых в течение некоторого периода времени или с использованием нескольких партий веществ может произойти изменения условий эксперимента. Для предотвращения смещения среднего уровня отклика точки экспериментального плана рекомендуется выбирать в случайной последовательности. Однако в некоторых случаях влияние внешних условий можно избежать путем подходящего разбиения матрицы на блоки по сырью или последовательности испытаний. Так, при наличии двух партий сырья матрицу 2^3 можно разбить на два блока таким образом, чтобы эффект сырья сказался на величине трехфакторного взаимодействия $x_1x_2x_3$ (табл. 3.12). В этом случае все линейные коэффициенты и парные взаимодействия будут освобождены от влияния неоднородности сырья.

Таблица 3.12

МПЭ ПФЭ 2^3

№ блока	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2\,x_3$	y
1	+	–	–	+	+	–	–	+	$y_1+\varepsilon$
	+	+	–	–	–	–	+	+	$y_2+\varepsilon$
	+	–	+	–	–	+	–	+	$y_3+\varepsilon$
	+	+	+	+	+	+	+	+	$y_4+\varepsilon$
2	+	–	–	–	+	+	+	–	y_5
	+	+	–	+	–	+	–	–	y_6
	+	–	+	+	–	–	+	–	y_7
	+	+	+	–	+	–	–	–	y_8

Различие в сырье рассматривают как новый фактор x_4. Матрица 2^3, представленная двумя блоками: первый блок $(x_1x_2x_3=+1)$ и второй блок $(x_1x_2x_3=-1)$, является полурепликой 2^{4-1} с определяющим контрастом $1=x_1x_2x_3x_4$.

Коэффициенты модели будут равны

$$b_0 = \frac{1}{8}[(y_1+\varepsilon)+(y_2+\varepsilon)+(y_3+\varepsilon)+(y_4+\varepsilon)+y_5+y_6+y_7+y_8]$$

$$b_0 \to \beta_0 + \frac{\varepsilon}{2}$$

$$b_1 = \frac{1}{8}[-(y_1+\varepsilon)+(y_2+\varepsilon)-(y_3+\varepsilon)+(y_4+\varepsilon)-y_5+y_6-y_7+y_8]$$

$$b_2 \to \beta_2$$

$$b_3 \to \beta_3$$

$$b_{12} \to \beta_{12}$$

$$b_{13} \to \beta_{13}$$

$$b_{23} \to \beta_{23}$$

$$b_{123} = \frac{1}{8}[(y_1+\varepsilon)+(y_2+\varepsilon)+(y_3+\varepsilon)+(y_4+\varepsilon)-y_5-y_6-y_7-y_8]$$

$$b_{123} \to \beta_{123} + \frac{\varepsilon}{2}.$$

Эффект сырья отразился на подсчете свободного члена b_0 и коэффициента b_{123}.

Аналогично можно разбить на два блока любой эксперимент $2^к$. Нужно только выбрать такой элемент взаимодействия, которым можно безболезненно пожертвовать. Обычно, при отсутствии априорных сведений выбирают взаимодействие самого высокого порядка $x_1x_2x_3$ для 2^3 , $x_1x_2x_3x_4$ для 2^4, $x_1x_2x_3x_4x_5$ для x_5 и т.д.

Если имеется четыре источника неоднородности, которые могут заметно исказить результаты эксперимента, то матрицу 2^4 разбивают на четыре блока таким образом, чтобы линейные эффекты были освобождены от влияния межблокового эффекта.

Схемы разбиения на блоки рототабельных планов второго порядка так же нужно выполнить с учетом того, чтобы различие между блоками не влияло на коэффициенты при линейных, перекрестных и квадратичных членах. Для того, чтобы не было искажения результатов эксперимента и результаты не зависели от блоковых эффектов число центральных точек должно быть выбрано подходящим образом.

3.7 Критерии оптимальности планов

Минимизацию чувствительности математической модели процесса к случайным факторам проведем с учетом дисперсионных оценок.

Предположим, что вид модели известен. Требуется найти лучшие оценки коэффициентов уравнения регрессии.

Если это уравнение линейное относительно параметров, то оценки уравнений регрессий определяются из системы нормативных

уравнений

$$X^T W X b = X^T W Y,$$

где $X \;= \{X_{ji}\}$ – матрица независимых переменных; $X^T = \{X_{ji}\}$ — матрица, получаемая транспонированием матрицы X ; $Y = Xb$ — уравнение регрессии в матричной форме; W — функция веса;
и тогда оценки коэффициентов

$$b = \left(X^T W X\right)^{-1}\left(X^T W Y\right).$$

Будем предполагать, что дисперсия выходной величины Y не зависит от независимых переменных X и поэтому примем функцию веса $W = 1$.

С учетом общего вида функциональной зависимости Y от X получаем следующие оценки:

$$b = \left(F^T F\right)^{-1} F^T Y = CZ, \qquad (3.18)$$

где $F = \begin{vmatrix} f_{11} & f_{12} & \cdots & f_{1k} \\ f_{21} & f_{22} & \cdots & f_{2k} \\ \dots & \dots & \dots & \dots \\ f_{n1} & f_{n2} & \cdots & f_{nk} \end{vmatrix} = \begin{vmatrix} f_1 \\ f_2 \\ \dots \\ f_n \end{vmatrix}$, $f_{ij}(x)$ — известные функции входных переменных $x_1, x_2, \dots, x_n$, т.е.

$$\eta_i = \sum_{j=1}^{k} b_j f_{ij}(x), \qquad (3.19)$$

$f = (f_{i1}, f_{i2} \dots f_{ik})$. Обозначим $Z = F^T Y$.

Если F^T - транспонированная матрица F, то

$$C = \left(F^T F\right)^{-1} \qquad (3.20)$$

— ковариационная матрица. Для линейной модели $C = \left(X^T X\right)^{-1}$.

В результате действия случайных факторов экспериментатор получает математическую модель отличную от (3.19): $Y_i = \eta_i + \varepsilon_i$.

Определим чувствительность решения (3.18) к небольшим изменениям выходной величины Y, а следовательно и Z за счет случайного фактора:

$$S_{\frac{b}{z}} = \frac{db}{dz} = C\frac{dz}{dz} \qquad (3.21)$$

Из выражения (3.21) следует, что чувствительность решения "b" по отношению к случайным факторам выражается ковариационной матрицей C.

План эксперимента выбирается с учетом минимизации

чувствительности решения (3.18) по отношению к случайным воздействиям. Однако в том виде, в каком она предоставлена в выражении (3.21) не представляется возможным дать количественную оценку чувствительности для разных планов F. Поэтому в качестве числовой характеристики чувствительности принимают некоторую из числовых характеристик ковариационной матрицы C. В зависимости от этого получают ряд *критериев оптимальности*.

1. План называется ***A* — *оптимальным***, если он минимизирует сумму квадратов главных полуосей эллипсоида рассеяния оценок. Известно, что при нормальном законе распределения результатов наблюдений, поверхностями постоянных значений функции плотности распределения $P\left(\frac{b}{\beta}\right)$ являются эллипсоиды. Характеристики эллипсоида полностью определяются элементами матрицы C. Для двухфакторного эксперимента дадим геометрическую интерпретацию (рис. 3.4).

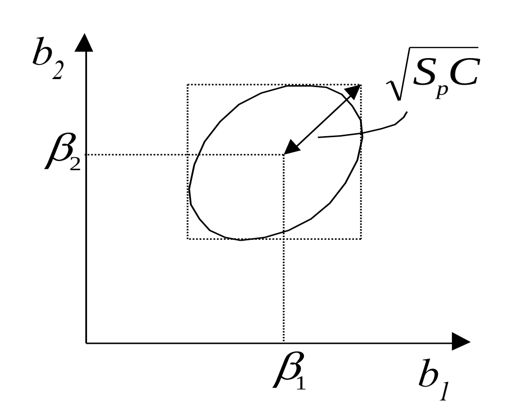


Рисунок 3.4 — Плоскость постоянных значений функции плотности при нормальном законе распределения

Через $S_p C$ обозначена сумма диагональных элементов матрицы C:

$$S_p C = \sum_{i=1}^{K} AC_{ii}$$

На языке матричной алгебры минимизация суммы квадратов главных полуосей эллипсоида означает минимизацию следа (суммы диагональных элементов) ковариационной матрицы уравнения регрессии. Еще можно встретить название минимизация средней дисперсии.

2. План называется ***D* — *оптимальным***, если он минимизирует величину определителя матрицы C. Это соответствует минимизации обобщенной дисперсии.

3. ***E* — оптимальный план** минимизирует максимальную ось эллипсоида рассеяния. Или еще можно назвать минимизацией максимального собственного числа матрицы C или минимизации максимального характеристического значения ковариационной матрицы уравнения регрессии.

4. ***G* — оптимальный план** обеспечивает наименьшую по всем планам максимальную величину дисперсии предсказанных значений функции отклика в области факторного эксперимента. Эти планы называют ***минимаксными***.

G — план можно определить из выражения

$$d(x) = f^T(x)\left(F^T F\right)^{-1} f(x_i)\sigma^2 \tag{3.22}$$

План будет G–оптимальным, если

$$d\left(x^*\right) = \min\max f^T(x)\left(F^T F\right)^{-1} f(x)\sigma^2 \tag{3.23}$$

Определитель матрицы C является наиболее полной числовой характеристикой чувствительности. D — оптимальным план учитывает все элементы матрицы C. Поэтому он лучше A и E — оптимальных планов. Что касается G — оптимальным, то он эквивалентен D — оптимальному плану. D — оптимальный план минимизирует объем эллипсоида рассеяния оценок коэффициентов модели.

Таким образом, в качестве числовой характеристики оптимальной оценки модели принимаем определитель матрицы C

$$S_{\frac{b}{z}} = \det C = \det\left(F^T F\right)^{-1} \tag{3.24}$$

Критерием выбора оптимального плана является минимизация чувствительности по всем возможным планам

$$S_{\frac{b}{z_{опт}}} = \min S_{\frac{b}{z}} = \min \det\left(F^T F\right)^{-1} \tag{3.25}$$

т.к. определитель матрицы $C = F^T F$ всегда положителен.

3.8 Синтез D–оптимальных тестирующих сигналов для идентификации динамических объектов

Метод идентификации статистических объектов называется ***пассивная идентификация***.

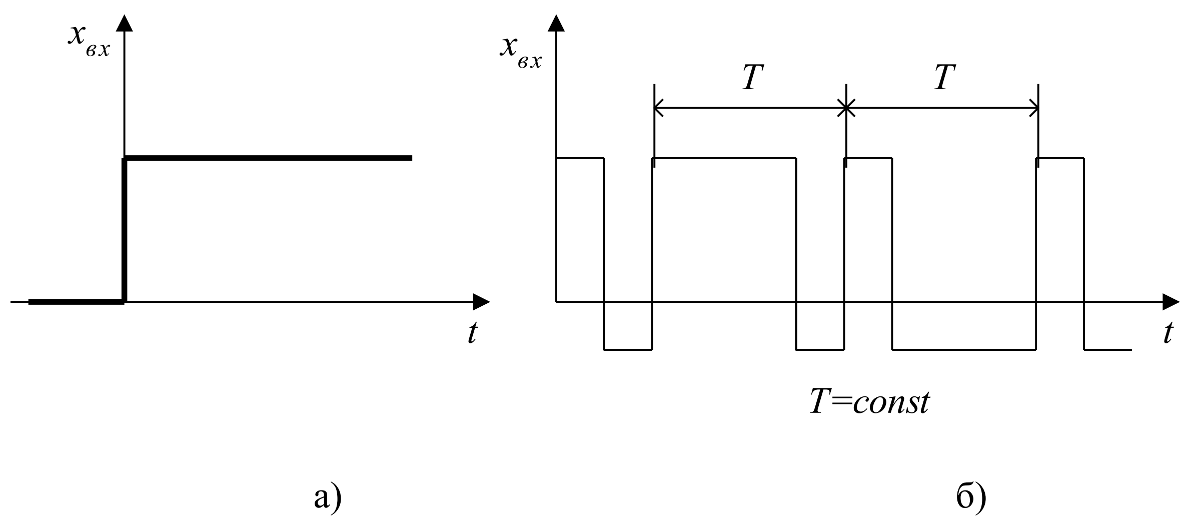
Методы идентификации динамических объектов — ***активная идентификация*** — представляет собой сложную задачу, т.к. их применение связаны с необходимостью нарушения технологического режима идентифицирующего объекта и имеют значительно больше возможностей.

Активная идентификация предполагает подачу на вход объекта специального тестирующего сигнала. При этом результаты идентификации, характеризуемые статистическими свойствами модели, полученной с помощью заданного метода оценивания (максимального правдоподобия, наименьших квадратов и т.п.), существенно зависят от структуры и параметров тестирующего сигнала. Выбор структуры параметров тестирующего сигнала представляют собой наиболее сложную часть общей задачи активной идентификации.

Кроме структуры и параметров, включающих всю необходимую информацию о тестирующем сигнале, как функции времени, на качество идентификации модели влияет также часть общего времени идентификации, состоящая из временной предварительной пассивной идентификации возмущений и помех, действующих на объект, и временного активного воздействия на объект, определяемого длиной тестирующего сигнала. Статистические свойства модели зависят также от моментов регистрации выходной координаты объекта относительно начала активного эксперимента идентификации.

Синтез (построение) D — оптимального тестирующего сигнала для линейных по параметрам динамических объектов сводится к следующему.

Существует несколько видов наиболее часто используемых сигналов: *ступенчатый*, *псевдослучайный* (рис. 3.15 а,б), *сигнал Дирека* (рис. 3.16).



а) б)

Рисунок 3.15 — Тестирующие сигналы: а) ступенчатый тестирующий сигнал; б) псевдослучайный тестирующий сигнал

Для идентификации динамических объектов применяют сигнал Дирека или $\delta(t)$ — дельта тестирующий сигнал (рис. 3.16).

$$\delta(t)=\begin{cases}\infty, t=\tau, \Delta\tau\to 0\\ 0, \quad t\neq\tau\end{cases}$$

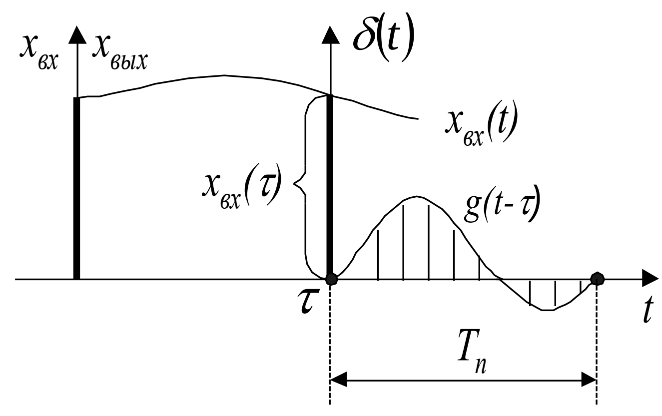


Рисунок 3. 16 — Тестирующий сигнал Дирека (*d(t)*– дельта сигнал)

Подача $\delta(t)$ на входе динамической системы дает реакцию системы в виде так называемой ***весовой функции*** $g(t-\tau)$. Причем длительность переходного процесса T_n.

Задача построения D— оптимального тестирующего сигнала для линейных по параметрам динамических объектов сводится к следующему:

1) находят D — оптимальный план для соответствующей регрессионной модели;

2) на основе полученного плана строят оптимальный тестирующий сигнал.

Предположим, что объект является линейным или допускает линеаризацию в пределах малых отклонений от некоторого статистического состояния. Пусть выход $\tilde{y}(t)$ измеряется дискретно с некоторым шагом Δt, а вектор шума $e^{\tau}=(e[0],e[\Delta t],e[2\Delta t]\ldots,e[n\Delta t])$. Входной сигнал $x(t)$ является кусочно-постоянной на интервале Δt функцией, тождественно равной нулю при отрицательном времени (до начала эксперимента).

В качестве модели линейного динамического объекта может быть принята регрессионная модель, описывающая поведение объекта в отклонениях от некоторого статического состояния:

$$Y[n\Delta t]=\sum_{m=0}^{l-1} g[m\Delta\tau]\cdot x[n\Delta t-(m+1)\Delta\tau]+e[n\Delta t], \qquad (3.26)$$

где $g[m\Delta\tau]$ — дискретная весовая функция (набор ординат); $x[n\Delta t-(m+1)\Delta\tau]$ — дискретное значение отклонения входного сигнала от их значений в исходном статическом состоянии; $Y[n\Delta t]$ —

дискретное значение отклонения выходного сигнала; $e[n\Delta t]$ — случайная помеха; $l\Delta\tau = T_n = x[n\Delta t-(m+1)\Delta\tau]$ — априорная оценка времени переходного процесса объекта.

Выражение (3.26) показывает, что единственным сигналом, которым можно варьировать в процессе эксперимента, являются дискретные значения входного сигнала. Эти сигналы и берутся в качестве факторов планирования эксперимента.

Таким образом, при $\Delta\tau = \Delta t$ факторами планирования будут

$$x[(n-i)\Delta t],\ i = 1,2,\ldots l \qquad (3.27)$$

Тогда, при условии, что $x(t)\equiv 0$ при $t < t_0$, пространство планирования X_0 будет ограничено системой неравенств

$$-1 \le x[(n-i)\Delta t] \le 1 \qquad (3.28)$$

Далее, известно, что для моделей типа (3.26) ортогональные планы, построенные с помощью ПФЭ или ДФЭ, являются одновременно и D— оптимальными. Поэтому для получения оптимального тестирующего сигнала проще использовать ПФЭ или ДФЭ. При этом следует иметь в виду, что в выражении (3.26) отсутствует свободный член. И эта особенность позволяет разбить матрицу ПФЭ на две, каждая из которых является D — оптимальным планом для данной модели. Разбиение матрицы на две делают с учетом того, что в каждой из полученных матриц один из факторов должен быть зафиксирован на одном из уровней +1 или –1.

Учитывая, что всего факторов l, то существует l симметричных пар эквивалентных D- оптимальных планов для модели (3.26), полученных на основе плана ПФЭ типа 2^l. Есть возможность найти и другие D–оптимальные планы.

Покажем *правило нахождения всех возможных* D — *оптимальных планов* для модели (3.26), матрицы которых имеют 2^{l-1} строк. Берется матрица ПФЭ 2^l. Строки этой матрицы формируются по две в каждой из $N = 2^{l-1}$ групп. В каждую группу включаются только симметричные строки матрицы ПФЭ. Затем все возможные D — оптимальные планы получаются путем включения одной из строк каждой группы в матрицу плана. Тогда общее количество возможных D — оптимальных планов для модели (3.26), полученных на основе ПФЭ типа 2^l будет равно 2^N.

Пример. Пусть необходимо найти четыре ординаты весовой функции, т.е. $l = 4$. Тогда ПФЭ будет иметь план $2^l = 2^4 = 16$ (табл. 3.13).

Таблица 3.13

МПЭ ПФЭ 2^4

Номер опыта табл.3.13	$x[(n-1)\Delta]$	$x[(n-2)\Delta]$	$x[(n-3)\Delta]$	$x[(n-4)\Delta]$
1	*-1*	*-1*	*-1*	*-1*
2	*+1*	*-1*	*-1*	*-1*
3	*-1*	*+1*	*-1*	*-1*
4	*+1*	*+1*	*-1*	*-1*
5	*-1*	*-1*	*+1*	*-1*
6	*+1*	*-1*	*+1*	*-1*
7	*-1*	*+1*	*+1*	*-1*
8	*+1*	*+1*	*+1*	*-1*
9	*-1*	*-1*	*-1*	*+1*
10	*+1*	*-1*	*-1*	*+1*
11	*-1*	*+1*	*-1*	*+1*
12	*+1*	*+1*	*-1*	*+1*
13	*-1*	*-1*	*+1*	*+1*
14	*+1*	*-1*	*+1*	*+1*
15	*-1*	*+1*	*+1*	*+1*
16	*+1*	*+1*	*+1*	*+1*

Перестраиваем матрицу 2^4, так чтобы в каждую группу были включены симметричные строки (табл. 3.14). D — оптимальные планы $2^{l-1} = 2^3 = 8$ получаются путем включения восьми строк, взятых по одной из каждой группы. Общее количество D — оптимальных планов будет равно $2^N = 2^8 = 256$, т.к. количество групп $N = 2^{4-1} = 8$.

Таблица 3.14

МПЭ ПФЭ 2^4 с симметричными строками

Номер опыта табл.3.14	*Номер группы*	$x[(n-1)\Delta]$	$x[(n-2)\Delta]$	$x[(n-3)\Delta]$	$x[(n-4)\Delta]$	*Номер опыта табл.3.13*
1	3	*-1*	*-1*	*-1*	*-1*	1
2		*+1*	*+1*	*+1*	*+1*	16
3	2	*+1*	*-1*	*-1*	*-1*	2
4		*-1*	*+1*	*+1*	*+1*	15
5	7	*-1*	*+1*	*-1*	*-1*	3
6		*+1*	*-1*	*+1*	*+1*	14
7	1	*+1*	*+1*	*-1*	*-1*	4
8		*-1*	*-1*	*+1*	*+1*	13
9	5	*-1*	*-1*	*+1*	*-1*	5
10		*+1*	*+1*	*-1*	*+1*	12
11	6	*+1*	*-1*	*+1*	*-1*	6
12		*-1*	*+1*	*-1*	*+1*	11
13	8	*-1*	*+1*	*+1*	*-1*	7
14		*+1*	*-1*	*-1*	*+1*	10

15	4	*-1*	*-1*	*-1*	*+1*	9
16		*+1*	*+1*	*+1*	*-1*	8

Теперь возникает задача построения самого D— оптимального тестирующего сигнала. Для этого существуют два пути.

Первый, наиболее простой, состоит в ***стыковке строк*** матрицы D — оптимального плана. Полученный таким путем тестирующий сигнал имеет максимальную длину, причем измерение выходной величины начинается через l — тактов, начиная с $(l+1)$.

Второй путь состоит в ***"сшивании" строк*** на γ тактов. Очевидно, что минимальная длина тестирующего сигнала достигается в том случае, если все строки матрицы "сшиваются" на $\gamma=(l-1)$ тактов. В этом случае измерение выходной величины осуществляется на каждом такте, начиная с $(l+1)$. Такое "сшивание" можно считать корректным, если получающийся при этом план является D — оптимальным. Для нашего случая полученный путем "сшивания" план должен быть представлен одним из 256 планов.

Используем метод "сшивания" для получения D — оптимального плана $2^{l-1}=2^{4-1}$ на основе матрицы таблицы 3.14.

На основе таблицы 3.15 получили минимальной длины D — оптимальный тестирующий сигнал *+1+1-1-1 -1-1+1-1 +1+1-1*.

Таблица 3.15

D— оптимальный план ДФЭ

Номер опыта табл.3.15	$x[(n-4)\Delta]$	$x[(n-3)\Delta]$	$x[(n-2)\Delta]$	$x[(n-1)\Delta]$	*Номер опыта табл.3.14*	*Номер группы*
1	*+1*	*+1*	*-1*	***-1***	8	1
2	*+1*	*-1*	*-1*	***-1***	15	4
3	*-1*	*-1*	*-1*	***-1***	1	3
4	*-1*	*-1*	*-1*	***+1***	3	2
5	*-1*	*-1*	*+1*	***-1***	5	7
6	*-1*	*+1*	*-1*	***+1***	11	6
7	*+1*	*-1*	*+1*	***+1***	10	5
8	*-1*	*+1*	*+1*	***-1***	13	8

В таблице 3.16 также представлен D–оптимальный план для нахождения четырех ординат весовой функции, однако длина тестирующего сигнала будет уже не оптимальной *+1+1-1-1 +1-1-1-1 -1+1-1+1*.

Таблица 3.16

D — оптимальный план ДФЭ

Номер опыта табл.4.16	$x[(n-4)\Delta]$	$x[(n-3)\Delta]$	$x[(n-2)\Delta]$	$x[(n-1)\Delta]$	*Номер опыта табл.4.15*	*Номер группы*
1	<u>**+1**</u>	<u>+1</u>	<u>-1</u>	<u>-1</u>	8	1
2	+1	-1	-1	**+1**	14	8
3	-1	-1	+1	**-1**	5	7
4	-1	+1	-1	**-1**	9	5
5	+1	-1	-1	**-1**	15	4
6	-1	-1	-1	**-1**	1	3
7	-1	-1	-1	**+1**	3	2
8	-1	+1	**-1**	**+1**	11	6

На рисунке 3.17 а, б) представлены сами тестирующие сигналы полученные соответственно по таблицам 3.15, 3.16.

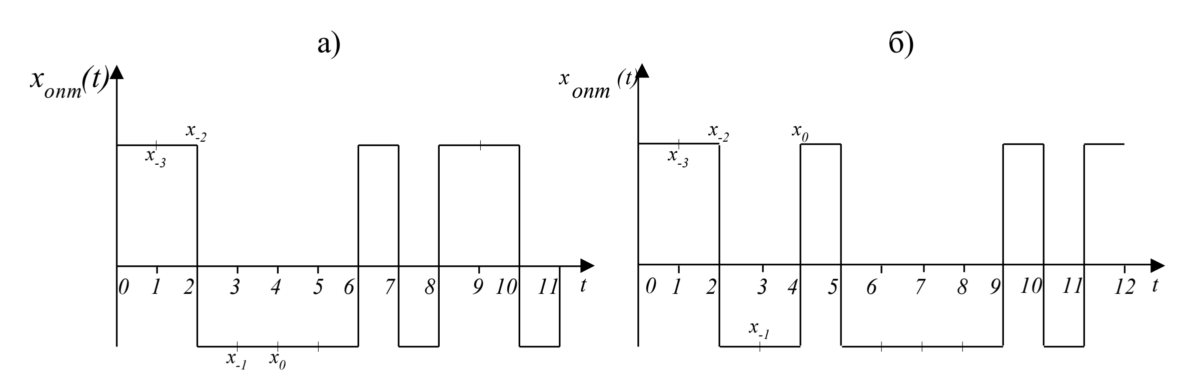


Рисунок 3.17 — Оптимальный тестирующий сигнал

С учетом полученного оптимального тестирующего сигнала, модель линейного динамического объекта для каждого опыта примет следующий вид:

$$Y[\Delta t] = g_0 x_0 + g_1 x_{-1} + g_2 x_{-2} + g_3 x_{-3};$$
$$Y[2\Delta t] = g_0 x_1 + g_1 x_0 + g_2 x_{-1} + g_3 x_{-2};$$
$$Y[3\Delta t] = g_0 x_2 + g_1 x_1 + g_2 x_0 + g_3 x_{-1};$$
$$Y[4\Delta t] = g_0 x_3 + g_1 x_2 + g_2 x_1 + g_3 x_0;$$
$$Y[5\Delta t] = g_0 x_4 + g_1 x_3 + g_2 x_2 + g_3 x_1;$$
$$Y[6\Delta t] = g_0 x_5 + g_1 x_4 + g_2 x_3 + g_3 x_2;$$
$$Y[7\Delta t] = g_0 x_6 + g_1 x_5 + g_2 x_4 + g_3 x_3;$$
$$Y[8\Delta t] = g_0 x_7 + g_1 x_6 + g_2 x_5 + g_3 x_4;$$
$$Y[9\Delta t] = g_0 x_8 + g_1 x_7 + g_2 x_6 + g_3 x_5.$$

Так как оптимальный тестирующий сигнал, построенный на основе таблицы 3.16, длиннее на один такт, то и реакций системы

будет получено на одну больше — не восемь, как для оптимального сигнала по таблице 3.15, а девять.

Контрольные вопросы

1. Каким образом при планировании эксперимента проводят выбор интервалов варьирования, выбор основного уровня плана эксперимента, выбор числа уровней?
2. Для каких видов моделей проводят полный факторный эксперимент? В чем заключается принцип рандомизации плана эксперимента?
3. Перечислите и сформулируйте свойства ПФЭ $2^к$.
4. Для каких моделей возможно проведение дробного факторного эксперимента?
5. Каким образом проводят планирование экспериментов при построении полной квадратичной модели?
6. Дайте определение понятия «звездные точки».
7. Какие критерии оптимальности планов Вы знаете?
8. В каких случаях проводиться разбиение матрицы $2^к$ на блоки? Каким образом изменятся оценки параметров модели?
9. Какие методы применяются для синтеза D — оптимальных тестирующих сигналов для идентификации динамических объектов? В чем они заключаются?

4. Экспериментирование

4.1 Статистические методы исследования объектов

4.1.1 Общие понятия и определения. Выборочная статистика

Статистические методы используют идеи и методы многомерной математической статистики. Формализация процессов здесь осуществляется эмпирическими методами [11, 13].

Основной задачей математической статистики является разработка методов получения научно обоснованных выводов о массовых явлениях и процессах из данных наблюдений или экспериментов.

В математической статистике вводится совсем новое и глубоко абстрактное понятие генеральной совокупности. Она представляет собой совокупность всех мыслимых наблюдений над случайной величиной при заданных условиях эксперимента. Если в результате эксперимента получены значения $x_1, x_2 \ldots x_n$, то эти величины образуют выборку, а число n называется ***объёмом выборки***.

Термин *выборочная статистика* обозначает некоторое числовое значение, подсчитанное на основе обработки данных выборки. Следует помнить, что выборка охватывает не всю генеральную совокупность, поэтому выборочные статистики являются случайными величинами. Выборочными статистиками являются : среднее значение, выборочная дисперсия и выборочный коэффициент корреляции случайной переменной величины x.

$f(x) = F^{'}(x)$ — это плотность вероятности или плотность случайной величины, где $F(x) = \int\limits_{-\infty}^{x} f(x)dx$ — функция распределения.

Выборочные статистики рассмотрим после функции распределения.

4.1.2 Функция распределения выборок

Функцией распределения $F(x)$ случайной величины ξ называют $F(x) = P(\xi \le x)$.

Полученные при эксперименте числа $x_1, x_2 \ldots x_n$ считают совокупностью значений некоторой конечнозначной величины x_n. При этом считают, что все эти числа являются различными элементами, независимо повторяются они или нет, т.е. каждый элемент выборки появляется лишь в результате одного измерения. Поэтому каждому элементу x_i приписывают вероятность $\frac{1}{n}$.

Полученное равномерное распределение величины x_n называют ***эмпирическим*** или ***выборочным распределением***. При увеличении объема выборки n распределение случайной величины x_n оказывается близким к распределению генеральной совокупности x. Это утверждение доказывается основной теоремой математической статистики.

Согласно этой теоремы: при $n \to \infty$ с вероятностью, равной единице, разность между функциями распределения случайных величин x_n и x стремится к нулю

$$\lim_{n\to\infty} \left| F_n(x) - F(x) \right| \to 0$$

Практически при достаточно большом объеме выборки функцию распределения генеральной совокупности можно приближенно заменить выборочной функцией распределения.

Вспомним, как вычисляется функция распределения. Для любого x_e функция распределения равна сумме вероятностей значений величины x_e не превосходящей x.

Считают, что все элементы выборки имеют одинаковую вероятность равную $\frac{1}{n}$. Поэтому функция распределения выборки в каждой точке равна числу элементов выборки, меньших, чем x_e, деленному на объем n.

Пример. Пусть выборка состоит из элементов (-1.8; -1; 0.3; 2; 5). Объем выборки $n = 5$. Тогда $F(x)$ каждого элемента: $F(x) = \frac{1}{n} = \frac{1}{5} = 0{,}2$. Определим функцию при $x = 1$ $F_5(1) = \frac{3}{5}$, при $x = 4$ $F_5(4) = \frac{4}{5} = 0{,}8$. Графическая иллюстрация примера приведена на рисунке 4.1.

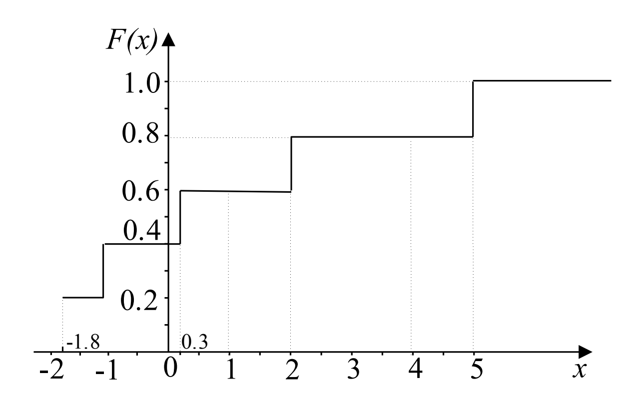


Рисунок 4.1 — Графическая иллюстрация примера

4.1.3 Числовые характеристики распределения

Среднее выборки. Математическое ожидание

Среднее выборки является наиболее эффективной выборочной статистикой. С помощью его оценивается математическое ожидание соответствующей случайной величины x, т.е. $\bar{x} \to M[x]$. Математическое ожидание для бесконечного ряда совокупности определяется следующим выражением: $M[x] = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx$, где $f(x)$– плотность вероятности. Математическое ожидание для конечного ряда совокупности – $M[x] = \sum\limits_{k=1}^{n} x_k f(x_k)$.

Случайная величина x_n конечнозначна и имеет равномерное распределение. Поэтому среднее выборки есть среднее арифметическое элементов выборки

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$

При достаточно большом n среднее может быть приближенно равным математическому ожиданию. При ограниченном числе опытов среднее выборки является случайной величиной, но все же оно дает более точное представление о математическом ожидании случайной величины x.

При моделировании идеальное значение любой задаваемой величины при расчетах интерпретируется как математическое ожидание.

Выборочная дисперсия

Выборочная дисперсия случайной величины x_n служит наилучшей оценкой дисперсии генеральной совокупности X. Выборочная дисперсия также является случайной величиной и определяется выражением:

$$D[x_n] = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n} = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n}.$$

Однако, учитывая, что на практике значение $\bar{x}$ не точно равно $M[x]$, т.е. $\bar{x} \cong M[x]$, дисперсия $D[x_n]$ дает, как правило, заниженную оценку рассеяния значений генеральной совокупности. Приведем пока без доказательства выражение:

$$D[x_n] = \frac{n-1}{n} D[x].$$

Из этого выражения видно, что истинный результат рассеяния значений генеральной совокупности $D[x]$ не совпадает с результатом рассеяния значений конечнозначной случайной величины $D[x_n]$. Значение $D[x_n]$ получается несколько меньше значения $D[x]$. Поэтому $D[x_n]$ называют ***смещенной оценкой дисперсии*** $D[x]$. В качестве дисперсии выборки рассматривают величину

$$S^2 = \frac{n}{n-1} D[x_n] \text{ или } S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{4.1}$$

Величина S^2 является несмещенной оценкой генеральной дисперсии

$$S^2 \to D[x]$$

Докажем несмещенность выборочной дисперсии S^2.

Преобразуем выражение (2.1)

$$(n-1)S^2 = \sum(x_i - \bar{x})^2 \tag{4.2}$$

Представим величину $(x_i - \bar{x})$ в следующем виде:

$$(x_i - \bar{x}) = (x_i - m_x) - (\bar{x} - m_x) \tag{4.3}$$

здесь $m_x = M[x]$.

Подставим правую часть выражения (4.3) в (4.2):

$$\begin{aligned}(n-1)S^2 &= \sum[(x_i - m_x) - (\bar{x} - m_x)]^2 = \\ &= \sum(x_i - m_x)^2 - 2\sum(x_i - m_x)(\bar{x} - m_x) + \sum(\bar{x} - m_x)^2 = \\ &= \sum(x_i - m_x)^2 - 2n(\bar{x} - m_x)^2 + n(\bar{x} - m_x)^2 = \\ &= \sum(x_i - m_x)^2 - n(\bar{x} - m_x)^2.\end{aligned} \tag{4.4}$$

Найдем математическое ожидание обеих частей выражения (4.4):

$$\begin{aligned}M\left[(n-1)S^2\right] &= M\left[\sum(x_i - m_x)^2 - n(\bar{x} - m_x)^2\right] = \\ &= M\left[\frac{n\sum(x_i - m_x)^2}{n}\right] - n\,D[\bar{x}] = \qquad \text{или} \\ &= n\,D[x] - n\frac{D[x]}{n} = D[x](n-1)\end{aligned}$$

$$(n-1)M\left[S^2\right] = (n-1)D[x].$$

Отсюда следует, что математическое ожидание выборочной дисперсии S^2 равно генеральной дисперсии $D[x]$. Значит S^2 является несмещенной оценкой генеральной дисперсии $D[x]$.

При моделировании идеальное значение любой задаваемой величины среднеквадратического отклонения при расчетах интерпретируется как дисперсия.

Выборочный коэффициент корреляции.

Понятие о корреляции

В математическом анализе зависимость между двумя величинами выражается функцией. Эта зависимость однозначно связывает независимую переменную с другой зависимой переменной. Эту зависимость называют ***функциональной***. Ее устанавливают с помощью строгих логических доказательств на основе определенных физических законов.

Намного сложнее дело обстоит с понятием зависимости между случайными величинами. Например, если при изменении случайной величины x изменится и другая случайная величина y, то нельзя сразу ответить, является ли это изменение результатом зависимости от x, или оно обязано лишь влиянию случайных факторов.

В природе существует причинно-следственная связь между случайными величинами. Эта связь особого рода, при которой с изменением одной величины меняется распределение другой. Такую связь называют ***стохастической связью***. Оценка силы этой связи представляет собой сложную задачу математической статистики. В общем виде эта задача до сих пор не решена. Однако существуют показатели, оценивающие те или иные стороны стохастической связи. При этом основным показателем является *коэффициент корреляции*.

Необходимым и достаточным условием корреляции между случайными величинами служит неравенство

$$M[(y-m_y)(x-m_x)] \neq 0,$$

где m_x, m_y — математическое ожидание величины x и y соответственно.

Величина $M[(y-m_y)(x-m_x)] = \sigma_{yx}^2$ носит название ***корреляционного момента*** или чаще называют ***ковариацией.***

На практике удобнее пользоваться безразмерным коэффициентом

$$\rho = \frac{M[(y-m_y)(x-m_x)]}{\sqrt{D_y \cdot D_x}} = \frac{\sigma_{yx}^2}{\sigma_y \sigma_x},$$

Этот коэффициент называется ***коэффициентом корреляции***. Для независимых величин коэффициент корреляции равен нулю.

Коэффициент корреляции характеризует не только наличие, но и силу стохастической связи между y и x. Максимальная корреляция равна $\rho = \pm 1$.

При $\rho > 0$ случайные величины y и x одновременно возрастают или убывают, а при $\rho < 0$ с возрастанием одной величины другая убывает или наоборот.

Коэффициент корреляции, как показатель зависимости случайных величин, имеет и слабую сторону: из равенства $\rho = 0$ еще не следует независимость величин y и x. Величины могут быть даже связаны функционально, а коэффициент корреляции может быть равен нулю. Это возможно при нелинейной зависимости. Однако, в тех случаях, когда заведомо известно, что между величинами существует линейная связь коэффициент корреляции является достаточно полным показателем.

Выборочный коэффициент корреляции r является лучшей оценкой коэффициента корреляции ρ (генерального коэффициента корреляции).

Пусть проведено m испытаний и при каждом из них фиксировались значения двух случайных величин. Тогда получается m- пар выборочных значений $(x_1, y_1);(x_2, y_2)\ldots(x_m, y_m)$. Эти точки можно рассматривать на плоскости и их расположение дает нам представление о силе корреляции.

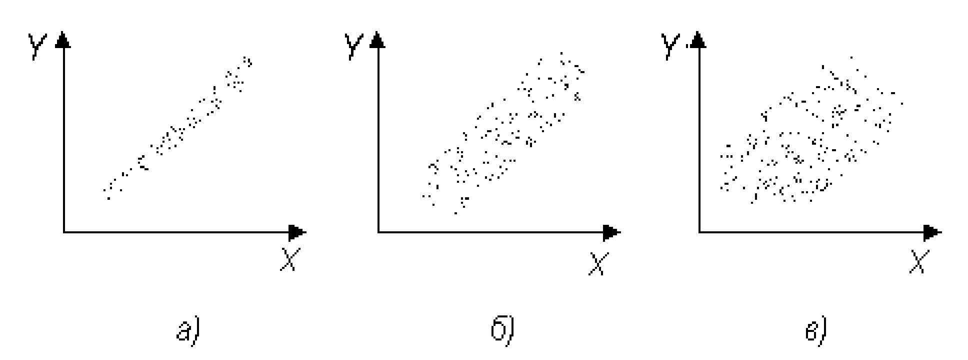


Рисунок 4.2 — Расположение экспериментальных точек на плоскости

На рис. 4.2 приведены примеры, соответствующие сильной корреляции, случай *а)*; слабой *б)* и полному ее отсутствию, случай *в)*.

Выборочный коэффициент корреляции находится аналогично, как и генеральный коэффициент корреляции.

Выборочный корреляционный момент равен

$$\rho_{yx}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y}),$$

а выборочный коэффициент корреляции:

$$r = \frac{\rho_{yx}}{S_x \cdot S_y},$$

где S_x^2, S_y^2 – выборочные дисперсии

$$S_x^2 = \frac{1}{m-1}\sum_{m=1}^{m}(x_i - \bar{x})^2; \; S_y^2 = \frac{1}{m-1}\sum_{m=1}^{m}(y_i - \bar{y})^2 .$$

Выборочный коэффициент корреляции с увеличением объема выборки m приближается к коэффициенту корреляции ρ.

Нужно помнить, что выборочный коэффициент корреляции дает оценку для коэффициента корреляции ρ только в тех случаях, когда совместное распределение двух случайных величин y и x является нормальным. При этом зависимость между y и x будет линейной. Однако, если только каждая случайная величина имеет нормальный закон распределения, то эти величины могут быть некоррелированными, но зависимы друг от друга. Только совместное нормальное распределение случайных величин дает возможность использовать выборочный коэффициент корреляции r для анализа стохастической связи между величинами.

4.1.3 Квантили распределения

Рассмотренные выборочные параметры: среднее, среднее квадратическое отклонение и коэффициент корреляции выборки являются приближенными оценками соответствующих генеральных параметров. Погрешность этих оценок будет тем меньше, чем больше объем выборки. Есть способы, с помощью которых можно оценить саму погрешность. Для этого переходят от точечных оценок параметров к оцениванию доверительных интервалов параметров. При получении интервальных оценок часто используют так называемые квантили.

Квантилем, отвечающий заданному уровню вероятности *P*, называют такое значение $x = x_p$, при котором функция распределения $F(x)$ принимает значение, равное *P*, т.е. $F[x_p] = P$, где *P* — заданный уровень вероятности.

Другими словами ***квантиль*** есть такое значение случайной величины x, при котором $F[x < x_p] = P$.

Вероятность *P*, задаваемая в процентах, дает название соответствующему квантилю, например $x_{0.4}$, называется 40%-ым квантилем.

Квантили стандартного нормального распределения (распределение с параметрами $m_x = 0, \sigma = 1$) обозначаются буквой u_p. Они сведены в соответствующих таблицах. Если $p < \frac{1}{2}$, то подбирая

такое x, для которого $\Phi(x)=\frac{1}{2}-p$ и находим $u_p=-x$. Если $p>\frac{1}{2}$, то подбираем такое x, для которого $\Phi(x)=p-\frac{1}{2}$ и тогда $u_p=x$. Например, 40 % квантиль будет равен $u_{0.4}=-0.25$, 85 %, квантиль $u_{0.85}=1.04$ и т.д.

Квантиль общего нормального распределения υ_p с параметрами m_x и σ выражается через квантиль u_p:

$$\upsilon_p=m_x+\sigma u_p. \qquad (4.5)$$

Если известны два квантиля случайной величины x_p и x_q, то

$$F[x_p\le x\le x_q]=q-p,$$

Понятие квантиля используется не только для нормального, но и для большинства встречающихся распределений.

Квантиль $x_{1/2}$ называется ***медианой распределения***. Если распределение случайной величины симметрично, то $x_{1/2}=M(x)$.

Например, распределение случайных ошибок симметрично. Поэтому для этого распределения можно использовать как математическое ожидание, так и медиану.

4.1.4 Получение интервальных оценок

Все выборочные параметры являются случайными величинами, а, следовательно, и их отклонения от генеральных параметров, т.е. погрешности, также будут случайными. Поэтому оценка этих погрешностей носит вероятностный характер. Можно лишь указать вероятность той или иной погрешности. Чтобы решить подобную задачу нужно найти вероятность того, что отклонения выборочного параметра ν от исследуемого генерального θ не превосходит по абсолютной величине некоторого заданного числа ε, т.е. находятся в пределах от $-\varepsilon$ до $+\varepsilon$. Обозначим отклонение через $\Delta\nu$. Поставленную задачу можно легко решить, если известна функция распределения $F(x)$ или плотность распределения $f(x)$ величины $\Delta\nu$:

$$P[|\Delta\nu|\le\varepsilon]=F(\varepsilon)-F(-\varepsilon)=\int_{-\varepsilon}^{\varepsilon}f(x)dx \qquad (4.6)$$

Распределение $\Delta\nu$ иногда удается точно определить по элементам выборки. В других случаях это распределение зависит только от объема выборки n и его можно вывести теоретически. Если бы при этом было известно математическое ожидание выборочного

параметра ν, то разность

$$M[\nu]-M[\Delta\nu]=M[\nu]-M[\nu-\theta]=M[\nu]-M[\nu]+\theta=\theta$$

дает точное значение генерального параметра. К сожалению $M[\nu]$, как правило, не известно.

Определим генеральный параметр. Находят по выборке одно значение ν_0 выборочного параметра ν и принимают его за приближенное значение генерального параметра θ. Затем, используя выражение (4.6) оценивают это приближение. Действительно, задаваясь некоторым положительным числом ε, находим вероятность того, что $|\Delta\nu|=|\nu-\theta|\le\varepsilon$. Но так как ν_0 есть одно из допустимых значений выборочного параметра ν, то вероятность неравенства $|\nu_0-\theta|\le\varepsilon$ также равна этой вероятности. Отсюда получаем формулу

$$P[|\nu_0-\theta|\le\varepsilon]=F(\varepsilon)-F(-\varepsilon)=\int_{-\varepsilon}^{\varepsilon}f(x)dx,$$

которая позволяет сравнивать найденное значение выборочного параметра с неизвестным генеральным параметром.

Неизвестный генеральный параметр можно представить в виде $\nu_0-\varepsilon\le\theta\le\nu_0+\varepsilon$.

Это неравенство отличается тем, что неизвестная величина θ, которая является неслучайной величиной, оценивается случайными границами, т.к. выборочный параметр является случайным.

Таким образом, любая статистическая оценка есть оценка вида

$$\nu'\le\theta\le\nu'',$$

где ν',ν'' некоторые случайные величины.

Придавая ν',ν'' конкретные значения, вычисляем вероятность соответствующей оценки.

В качестве границ ν',ν'' наиболее удобно брать квантили случайной величины ν. При обработке наблюдений для оценок генерального параметра берут симметричные квантили. В этом случае вероятности *P* соответствует оценка $\nu_{\frac{1-p}{2}}\le\theta\le\nu_{\frac{1+p}{2}}$,

Рассмотрим, как пользоваться оценками на практике. Для этого дадим следующие определения.

Событие называется ***абсолютно достоверным***, если оно появляется при любом осуществлении комплекса основных факторов. Абсолютную достоверность нельзя установить никакой самой длительной проверкой. Ее можно вывести лишь теоретически, путем логических умозаключений. Сюда относят обычно математические истины.

Большинство же привычных достоверных событий при

рассмотрении не является абсолютно достоверным. Например, нельзя считать абсолютно достоверным тот факт, что подброшенная монета упадет гербом или числом, т.к. у монеты есть и другие состояния равновесия - это ребро. Однако, в данном случае можем утверждать, что монета упадет либо гербом, либо числом. Такая достоверность называется ***практической достоверностью***.

Использования принципа практической достоверности позволяет не доводить вероятность оценки до единицы. Принимаемый при этом уровень вероятности называется ***доверительной вероятностью***. В зависимости от конкретных обстоятельств в качестве доверительной вероятности берут обычно значения: $0.95 \div 0.99$; реже 0.90;0.999.

Соответствующие доверительной вероятности квантильные границы называются ***доверительными границами***, а образуемый ими интервал — ***доверительным интервалом*** или еще называют ***доверительной оценкой*** или ***интервальной оценкой***.

Величина $\alpha = 1 - p$ называется ***уровнем значимости***. Уровень значимости соответствует практически невозможному событию. Уровень значимости и уровень достоверности в сумме дают единицу. Обычно значения уровней значимости берутся в пределах: $0.05 \div 0.01$ и реже 0.1; 0.001.

4.2 Генераторы случайных чисел. Псевдослучайные числа

При построении стохастических имитационных моделей необходимо обеспечить возможность учета стохастических воздействий. Результаты статистического моделирования существенно зависят от качества выбранных исходных последовательностей случайных чисел. На практике используется три основных способа генерации случайных чисел: аппаратный (физический), табличный (файловый) и алгоритмический (программный) [6, 13].

Определим основные требования, которым должен удовлетворять идеальный генератор. То есть псевдослучайные последовательности чисел должны:

1) состоять из квазиравномерно распределенных чисел;
2) содержать статистически независимые числа;
3) быть воспроизводимыми;
4) программа генерации должна иметь наименьшие затраты машинного времени и занимать минимальный объем памяти.

4.2.1 Методы получения псевдослучайных чисел

Метод средних квадратов

Одной из первых алгоритмических процедур получения псевдослучайных чисел был метод средних квадратов, который заключается в следующем. Например, для восьмиразрядного числа (половина равна 4 разрядам) берется начальное значение:
$x_0 = 0.2152$, возводится в квадрат:
$(x_0)^2 = 0.04631104$, берется "среднее" число:
$x_1 = 0.6311$ и снова возводится в квадрат:
$(x_1)^2 = 0.39828721$,
$x_2 = 0.8287$ и т. д.

Однако этот метод работает неудовлетворительно: существует наличие корреляции между числами, а иногда и отсутствие характера случайности.

Конгруэнтные методы

Самое широкое применение при моделировании на ЭВМ получили конгруэнтные методы генерации псевдослучайных последовательностей, в основе которых лежит фундаментальное понятие *конгруэнтности*.

Два целых числа α и β конгруэнтны (сравнимы) по модулю m, где m — целое число, тогда и только тогда, когда существует такое целое число k, что

$$\alpha - \beta = km,$$

то есть разность $\alpha - \beta$ делится на m, и если числа α и β дают одинаковые остатки деления на абсолютную величину числа m. Например:

$$1984 \equiv 4 \pmod{10}, 5008 \equiv 8 \pmod{10^3} \text{ и т.д.}$$

Величина m берется равной длине машинного слова $m = 2^b$, где b — число бит в машинном слове.

Конгруэнтный метод представляет собой арифметическую процедуру для генерирования конечной последовательности равномерно распределенных чисел. С использованием нескольких рекуррентных формул было построено множество конгруэнтных алгоритмов. Среди них наиболее известными являются мультипликативный, смешанный и аддитивный методы.

Мультипликативный конгруэнтный метод

Основная формула этого метода имеет вид:

$$x_{i+1} = ax_i \pmod m,$$

где a и m — неотрицательные целые числа.

Согласно этому выражению, должны взять последнее случайное число x_i, умножить его на постоянный коэффициент a и взять модуль полученного числа по m, то есть разделить на ax_i, и остаток считать

как x_{i+1}. Поэтому для генерирования последовательности чисел x_i необходимо начальное значение x_0, множитель a и модуль m.
Любой генератор псевдослучайных чисел может дать лишь конечное множество чисел, после которого последовательность будет повторяться. Период, или длина последовательности L, зависит от ЭВМ и от выбранного модуля, а статистические свойства зависят от выбранного начального значения и множителя. Таким образом, выбирать a, x_0 и m следует так, чтобы обеспечить максимальный период и минимальную корреляцию между генерируемыми числами. Правильный выбор модуля не зависит от системы счисления. Обычно берут m равным длине машинного слова:

– для двоичной ЭВМ $m=2^b$, где b - число двоичных цифр (бит) в машинном слове;

– для десятичной ЭВМ $m=10^d$, где d - число десятичных цифр в машинном слове.

Тогда, если правильно выбрать a и x_0 , то максимальный период будет равен

$L=2^{b-2}=m/4$ когда $b>2$ (двоичная система счисления);

$L=5 \cdot 10^{d-2}=m/20$ когда $d>2$ (десятичная система счисления).

Для двоичной системы $a = 8T \pm 3$, где T может быть любым целым положительным числом, x_0 — любое положительное, но нечетное число.

Для десятичной системы $a = 200T \pm Q$, где T любое целое положительное число, Q может принимать одно из следующих значений ± (3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61); x_0 — любое положительное число, которое не делится на 2 и на 5.

Процедура вычисления случайных чисел между 0 и 1:

1) выбрать в качестве x_0 произвольное нечетное число;
2) вычислить коэффициент a;
3) найти произведение ax_0, содержащее не более $2b$ значащих разрядов;
4) взять $b(d)$ младших разрядов в качестве первого члена последовательности x_1, а остальные отбросить;
5) определить дробь $x_1 = x_1 / 2^b$ ($x_1 = x_1 / 10^d$) из интервала (0, 1);
6) приравнять $x_0 = x_1$;
7) вернуться к пункту 3.

Пример. Надо получить последовательность чисел для случая $b=4$.

1. Выбираем $x_0=0111(7)$. Берем $T=1$.
2. $a=8T-3=8-3=5$, $a=0101$.
3. $ax_0=(0101)(0111)=00100011$.
4. $x_1=0011$.

5. Проверяем дробь интервала (0,1) $x_1=x_1/2^4=3/16=0{,}1875$. В результате получаем:

1) $ax_0=(0101)(0111)=00100011,\ x_1=0011,\ x_1=3/16=0{,}1875;$

2) $ax_1=(0101)(0011)=00001111,\ x_2=1111,\ x_2=15/16=0{,}9375;$

3) $ax_2=(0101)(1111)=01001011,\ x_3=1011,\ x_3=11/16=0{,}6875;$

4) $ax_3=(0101)(1011)=00110111,\ x_4=0111,\ x_4=7/16=0{,}4375.$

Таким образом, генерируемая последовательность чисел x_i, $i = 0, 1, ..., N$, представляет собой равномерно распределенные на интервале (0, 1) случайные числа.

Для дискретной последовательности используют 2^n случайных чисел на интервале (0,1). Поэтому закон распределения такой дискретной последовательности называют ***квазиравномерным распределением***. При этом случайная величина принимает значения:

$x_i = \dfrac{i}{2^n - 1}$ с вероятностью $p_i = \dfrac{1}{2^n}$, где $i = 0, 1, ..., 2^n - 1$.

Тогда математическое ожидание и дисперсия дискретной последовательности соответственно равны:

$$M(\xi) = \sum_{i=0}^{2^n-1} \frac{i}{2^n-1} \cdot \frac{1}{2^n} = \frac{1}{(2^n-1)\cdot 2^n} \sum_{i=0}^{2^n-1} i = \frac{(2^n-1)\cdot 2^n}{(2^n-1)\cdot 2^n \cdot 2} = \frac{1}{2},$$

$$D(\xi) = \sum_{i=0}^{2^n-1} \frac{1}{2^n}\left[\frac{i}{2^n-1} - \frac{1}{2}\right]^2 = .$$

$$= \frac{1}{2^n} \sum_{i=0}^{2^n-1}\left[\frac{i^2}{(2^n-1)^2} - \frac{i}{2^n-1} + \frac{1}{4}\right] = \frac{1}{12} \cdot \frac{2^n+1}{2^n-1}$$

Для равномерной случайной последовательности этого же интервала (0,1) имеем:

$$M(\xi) = \int_0^1 x f(x) dx = \frac{1}{2},$$ где $f(x)$ - функция плотности,

$$f(x) = \begin{cases} 1, 0 \le x \le 1 \\ 0 \end{cases};$$

$$D(\xi) = \int_0^1 \left[x - M(\xi)\right]^2 f(x) dx = \frac{1}{12}.$$

При достаточно больших n свойства законов распределения квазиравномерной последовательности и равномерной последовательности совпадают. Кроме того, для получения значений x используется алгоритмический способ. Поэтому такие последовательности по своей сути являются детерминированными и называются ***псевдослучайными***.

Смешанный метод

Работа этих генераторов основана на использовании формулы:

$$x_{i+1} = (a\,x_i + C)\ (\mathrm{mod}\ m),\ C > 0\,.$$

С вычислительной точки зрения смешанный метод генерации последовательности неотрицательных целых чисел сложнее мультипликативного на одну операцию сложения, но при этом возможность выбора дополнительного параметра позволяет уменьшить возможную корреляцию получаемых чисел. Однако экспериментальная проверка качества генерируемой последовательности чисел на основе этой формулы является сложнее.

4.2.2 Проверка качества квазиравномерной последовательности псевдослучайных чисел

Эффективность моделирования на ЭВМ существенно зависит от качества исходных последовательностей псевдослучайных чисел. Поэтому, прежде чем приступить к реализации моделирующих алгоритмов на ЭВМ, необходимо выполнить тестирование случайных чисел, которое включает в себя проверку равномерности, проверку стохастичности и проверку независимости.

Проверка равномерности выполняется по гистограмме. Для проверки выдвигается гипотеза о равномерности распределения чисел в интервале (0, 1). Затем этот интервал разбивается на m равных частей. Это означает, что при генерации последовательности $\{ x_i \}$ каждое из чисел x_i с вероятностью

$$P_j = \frac{1}{k}, j = \overline{1,k}$$

попадет в один из частей интервала. Всего в каждый j-ый подинтервал попадет N_j чисел, причем общее количество чисел:

$$N = \sum_{j=1}^{k} N_j\,.$$

Относительная частота попадания случайных чисел в каждый из подинтервалов равна N_j / N. На рисунке 4.3 показан пример гистограммы распределения чисел.

Пунктирная линия соответствует теоретическому значению P_j, а сплошная — экспериментальному N_j / N. Очевидно, что при большом N экспериментальная гистограмма приблизится к теоретической.

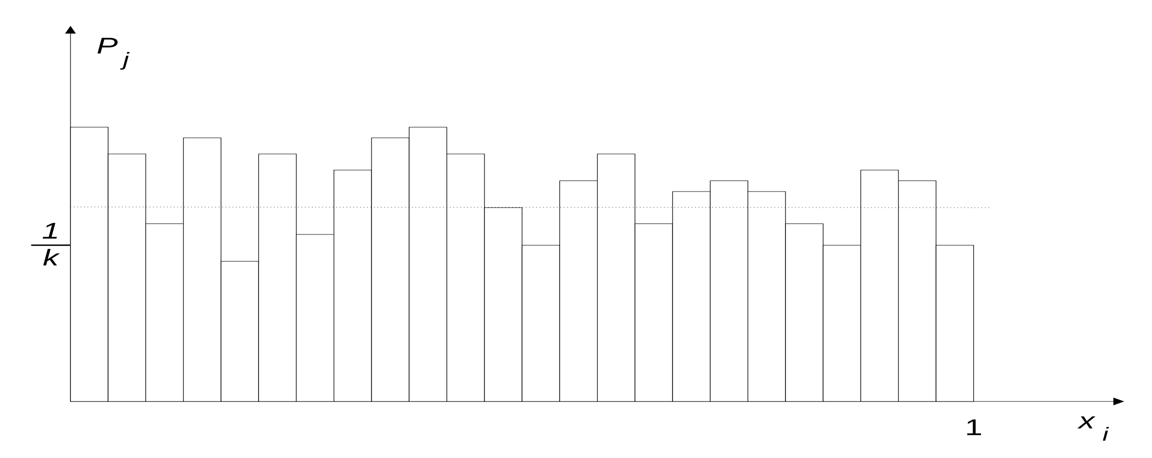


Рисунок 4.3 — Гистограмма распределения чисел

Оценка степени приближения может быть проведена с использованием критериев согласия (Колмогорова, Пирсона, Стьюдента, Фишера и т.п.). На практике обычно принимается $k = 20 \div 50$, $N = (10^2 \div 10^3)\, k$.

Проверка стохастичности

Среди наиболее часто используемых тестов являются *интервальный тест* и *автокорреляционные тесты*.

Интервальный тест проводится методами комбинаций и серий. Сущность метода комбинаций сводится к определению закона распределения длин участков между единицами (нулями) или закона распределения появления числа единиц (нулей) в n-разрядном двоичном числе x_i. На практике длину последовательности N берут достаточно большой и проверяют все n разрядов или только l старших разрядов числа x_i.

Теоретически закон появления j единиц в l разрядах двоичного числа x_i описывается биноминальным законом распределения:

$$P\,(j,l\,) = C_l^{\,j}\,P^{\,l}(1),$$

где $P\,(j,l\,)$ — вероятность появления j единиц в l разрядах числа x_i, $P\,(1) = =P\,(0) = 0.5$ — вероятность появления единицы (нуля) в любом разряде числа x_i; C_l^j — количество комбинаций появления единиц (нулей) в l разрядах:

$$C_l^j = \frac{l!}{j!\cdot(l-j)!}\ .$$

Тогда, если выбрать длину выборки N фиксированной, то теоретически ожидаемое число появления случайных чисел x_i с j единицами в проверяемых l разрядах будет равна

$$\eta_j = N\,C_l^{\,j}\,P^{\,l}\,(1)\ .$$

После нахождения теоретических и экспериментальных вероятностей $P(j,l)$ или чисел η_j при различных значениях $l \le n$

гипотеза о стохастичности проверяется с использованием критерия согласия.

Проверка независимости

Проверка независимости элементов последовательности $\{x_i\}$ проводится на основе вычисления корреляционного момента. Случайные числа ξ и η называются ***независимыми***, если закон распределения каждого из них не зависит от того, какое значение приняло другое. Для независимых величин корреляционный момент

$$K_{\xi\eta} = \sum_i \sum_j [x_i - M(\xi)][y_i - M(\eta)]p_{ij},$$

и коэффициент корреляции равны нулю

$$\overline{P}_{\xi\eta} = \frac{k_{\xi\eta}}{\sigma_x \cdot \sigma_y},$$

где p_{ij} — вероятность того, что (ξ, η) примет значение (x_i, y_i); σ_x, σ_y — среднеквадратические отклонения величин ξ и η.

Таким образом, независимость элементов последовательности $\{x_i\}$ может быть проведена путем введения в рассмотрение последовательности $\{y_i\}$, для которых $y_i = x_{i+\tau}$, где τ — величина сдвига последовательностей.

При проведении оценок корреляции на ЭВМ удобно для вычисления использовать следующие выражения:

$$\overline{P}_{\xi\eta} = \frac{\left[\frac{1}{N-\tau}\sum_{i=1}^{N-\tau} x_i x_{i+\tau} - \frac{1}{(N-\tau)^2}\sum_{i=1}^{N-\tau} x_i \sum_{i=1}^{N-\tau} x_{i+\tau}\right]}{\sqrt{D(x_i)D(x_{i+\tau})}}$$

$$D(x_i) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} x_i^2 - \frac{1}{(N-\tau)^2}\left(\sum_{i=1}^{N-\tau} x_i\right)^2$$

$$D(x_{i+\tau}) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} x_{i+\tau}^2 - \frac{1}{(N-\tau)^2}\left(\sum_{i=1}^{N-\tau} x_{i+\tau}\right)^2$$

При любом $\tau \neq 0$ для достаточно больших N с доверительной вероятностью β справедливо соотношение:

$$\left|\overline{P}_{\xi\eta}(\tau)\right| \leq \beta\sqrt{\frac{1}{N}},$$

где обычно принимает значение в интервале $\beta = 0{,}9 \div 0{,}99$.

Если найденное эмпирическое значение $\overline{P}_{\xi\eta}(\tau)$ находится в

указанных пределах, то с вероятностью β можно утверждать, что полученная последовательность чисел x_i удовлетворяет гипотезе корреляционной независимости.

4.3 Моделирование случайных сигналов и процессов

4.3.1 Формирование возможных значений случайных величин с заданным законом распределения

Исходным материалом служат случайные величины с равномерным распределением в интервале (0, 1):

$$f(x)=\begin{cases}1, & 0\le x\le 1\\ 0\end{cases}$$, где $f(x)$ — функция распределения.

Есть простой метод преобразования равномерного распределения в требуемое. Его называют ***прямым методом***. Он заключается в преобразовании случайных величин. Пусть переменная y является функцией от x:

$$y = \varphi(x).$$

Тогда для каждого значения x однозначно определяется y.

Пусть x — случайная величина, тогда y будет так же случайной величиной. И если x имеет функцию распределения $f_1(x)$, то y будет иметь так же функцию распределения $f_2(y)$, но зависящую уже от $f_1(x)$ и $\varphi(x)$.

Известно, что для взаимно однозначного соответствия между x и y имеет место следующее равенство:

$$f_1(x)\,dx = f_2(y)\,dy, \qquad (3.1)\ (4.7)$$

то есть вероятность того, что случайная величина x находится в интервале между x_i и x_i+dx, равна вероятности того, что преобразованная случайная величина y будет находиться в интервале y_i и y_i+dy.

Последнее выражение можно изобразить графически (рис. 4.4).

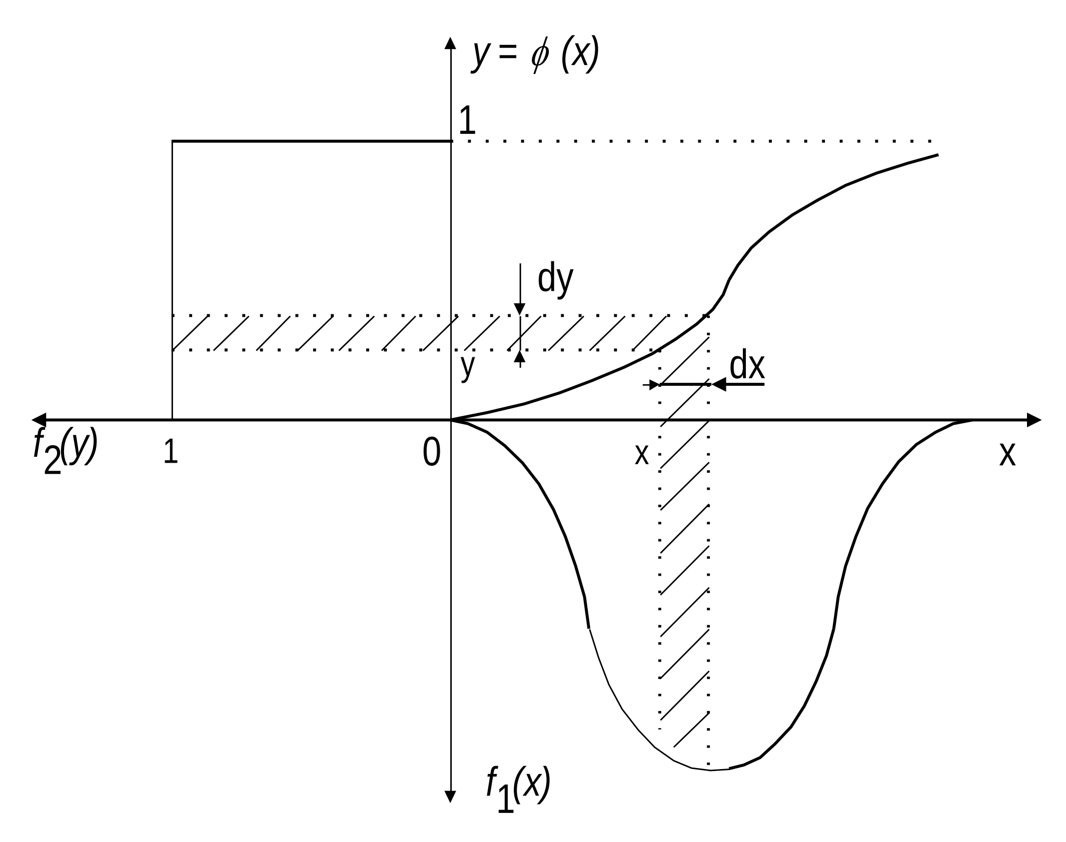


Рисунок 4.4 — Прямой метод преобразования

Согласно выражению (4.7) находится требуемое распределение случайной величины x:

$$f_2(y) = f_1(x)\frac{1}{\frac{dy}{dx}} = f_1(x)\frac{1}{|\varphi'(x)|}\ .$$

Требуемое распределение f_2 (y) является функцией от распределения случайной величины x и производной от функции φ (x). Абсолютное значение производной принято потому, что функция распределения не может принимать отрицательных значений. Функцию $y = \varphi(x)$ можно найти из (3.1):

$$\varphi'(x) = \frac{f_1(x)}{f_2(y)} \text{ или } \varphi(x) = \int_{-\infty}^{x} \frac{f_1(x)}{f_2(y)}dx \qquad (4.8)$$

Пусть f_2 (y) = $const$. Тогда имеем равномерное распределение, и выражение (4.8) принимает вид:

$$\varphi(x) = \int_{-\infty}^{x} f_1(x)dx \text{ или } \varphi(x) = F_1(x).$$

Функция φ (x) для данного случая совпадает с интегральной функцией распределения F_1 (x). Это выражение можно использовать для преобразования случайных чисел, то есть для реализации

некоторой операции над числом y_i, имеющим равномерный закон распределения, и преобразование его в число x_i, имеющее заданный закон распределения:

$$y_i = \int_{-\infty}^{x_i} f_1(x)dx.$$

Эта формула является рабочей. Она может быть использована для практических целей.

Пример. Необходимо получить случайные числа x_i с показательным законом распределения:

$$f(x) = \lambda e^{-\lambda x} \quad (x > 0). \qquad (4.9)$$

Согласно этому выражению имеем:

$$y_i = \lambda \int_{-\infty}^{x_i} e^{-\lambda x} dx.$$

Здесь y_i — случайная величина с равномерным законом распределения в интервале (0, 1).

После вычисления интеграла получим:

$$y_i = 1 - e^{-\lambda x_i} \text{ или } x_i = -\frac{1}{\lambda}\ln(1 - y_i).$$

И далее, учитывая, что случайная величина $y_i = 1 – y$ имеет так же случайное распределение в интервале (0, 1), запишем

$$x_i = -\frac{1}{\lambda}\ln y_i. \qquad (4.10)$$

Однако для многих практических случаев, например для нормального распределения, уравнение (4.9) не имеет точного решения. Кроме того, даже простое на первый взгляд выражение оказывается неудобным для решения его на ЭВМ. Поэтому *на практике* обычно пользуются приближенными приемами преобразования случайных чисел, основанными на *кусочной аппроксимации функции плотности* (рис.4.5).

4.3.2 Алгоритм формирование случайных величин

Пусть требуется получить последовательность случайных чисел $\{ y_i \}$ с функцией плотности $f_\eta (y)$, возможные значения которой лежат в интервале (a,$в$). Представим f_η (y) в виде кусочно-постоянной функции. Для этого разобьем интервал (a, $в$) на m интервалов и будем считать, что f_η (y) на каждом интервале постоянна. Чтобы аппроксимировать f_η (y) наиболее удобным способом для практических целей, целесообразно разбить (a, $в$) на интервалы так, чтобы вероятность попадания случайной величины η в любой

интервал (a_k, a_{k+1}) была постоянной, то есть не зависела от номера интервала k. Тогда для вычисления a_k можно использовать выражение

$$\int_{a_k}^{a_{k+1}} f_{\eta}(y)dy = \frac{1}{m} \qquad (4.11)$$

Случайную величину η можно представить в виде $\eta = a_k + \eta_k$, где a_k — абсцисса левой границы k-го интервала; η_k — случайная величина, возможные значения которой располагаются равномерно внутри k-го интервала.

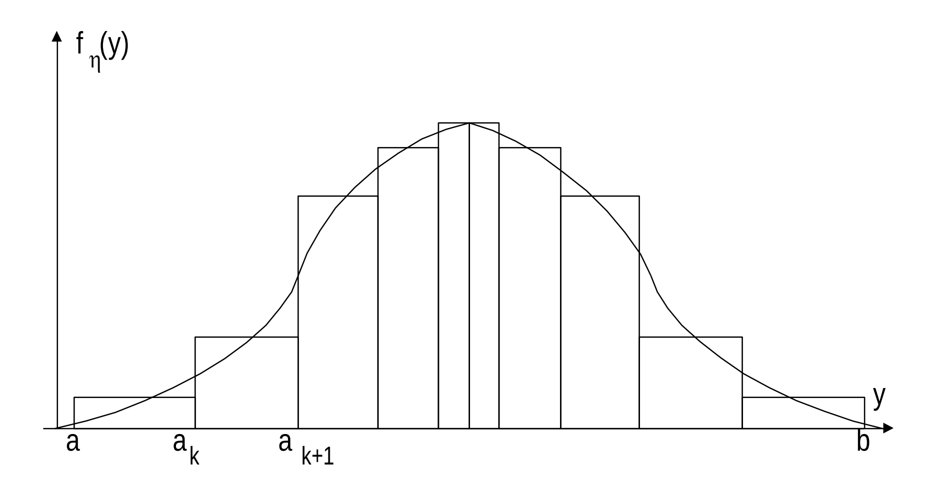


Рисунок 4.5 — Кусочная аппроксимация функции плотности

Алгоритм этого способа получения случайных чисел сводится к следующим действиям:

1) генерируется случайное равномерно распределенное число x_i из интервала (0, 1);

2) с помощью этого числа случайным образом выбирается интервал (a_k, a_{k+1});

3) генерируется число x_{i+1} и масштабируется с целью приведения его к интервалу (a_k, a_{k+1}), то есть помножается на коэффициент $(a_{k+1} - a_k)\ x_{i+1}$;

4) вычисляется случайное число $y_j = a_k + (a_{k+1} - a_k)\ x_{i+1}$ с требуемым законом распределения.

Для выбора интервала (a_k, a_{k+1}) с помощью случайного числа x_i строят таблицу (формируют массив), в которую предварительно помещают номера интервалов k и значения коэффициента масштабирования, определенные из соотношения (4.11) для приведения числа к интервалу (a, $в$). Получив случайное число x_i, с помощью этой таблицы определяют абсциссу левой границы a_k и коэффициент масштабирования ($a_{k+1} - a_k$).

Достоинство этого приближенного способа преобразования

случайных чисел состоит в том, что при реализации на ЭВМ требуется небольшое количество операций для получения каждого случайного числа, так как операция масштабирования выполняется только один раз перед моделированием, и количество операций не зависит от точности аппроксимации, то есть количества интервалов m.

Можно привести еще способы преобразования равномерно распределенных случайных чисел { x_i } в последовательность с заданным законом распределения { y_j } на основе предельной теоремы теории вероятностей.

4.4 Центральная предельная теорема

Если ξ_1, ξ_2, ..., ξ_N — независимые одинаково распределенные случайные величины, имеющие математическое ожидание a и дисперсию σ^2, то при $N \to \infty$ закон распределения суммы случайных чисел $\sum_{i=1}^{N} x_i$ неограниченно приближается к нормальному закону.

Рассмотрим применение предельной теоремы на примере.

Пусть требуется получить последовательность случайных чисел { y_j }, имеющих нормальное распределение с математическим ожиданием a и средним квадратичным отклонением σ.

$$f_\eta(y) = \frac{e^{-\frac{(y-a)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma}}$$

Случайные числа y_i формируются в виде сумм последовательностей случайных чисел { x_i }, имеющих равномерное распределение на интервале (0, 1). Тогда можно использовать центральную теорему.

Если независимые одинаково распределенные случайные величины x_1, x_2, ..., x_N имеют каждая математическое ожидание a_1 и среднее квадратичное отклонение σ_1, то сумма $\sum_{i=1}^{N} x_i$ асимптотически нормальна с математическим ожиданием $a = Na_1$ и $\sigma = \sigma_1\sqrt{N}$.

Расчеты показывают, что сумма $\sum_{i=1}^{N} x_i$ имеет распределение, близкое к нормальному, уже при $N = 8 \div 12$.

4.5 Правило автоматической остановки имитационного эксперимента

Первый способ. Простейший способ состоит в задании требуемого количества реализаций N. Однако такой подход является грубым из-за того, что на этапе тактического планирования неизвестны распределения выходных переменных.

Второй способ состоит в задании доверительных интервалов для выходных переменных. Остановка прогона машинной модели происходит при достижении заданного доверительного интервала.

Третий способ выполняется путем двухэтапного проведения прогона, когда сначала делается пробный прогон из N_1 реализаций, позволяющий оценить необходимое количество реализаций N. Если $N_1 > N$, то прогон можно закончить, в противном случае необходимо набрать еще $(N - N_1)$ реализаций.

Четвертый способ выполняется путем последовательного анализа для определения минимально необходимого количества реализаций N, которое рассматривается при этом как случайная величина, зависящая от N-1 предыдущих реализаций при достижении нормального закона распределения. Такой подход позволяет проводить анализ результатов эксперимента на базе характеристик выборок.

Рассмотрим четвертый способ на примере двухфакторной линейной (регрессионной) модели (теоретически данный вопрос рассматривался в п.п. 5.2.2).

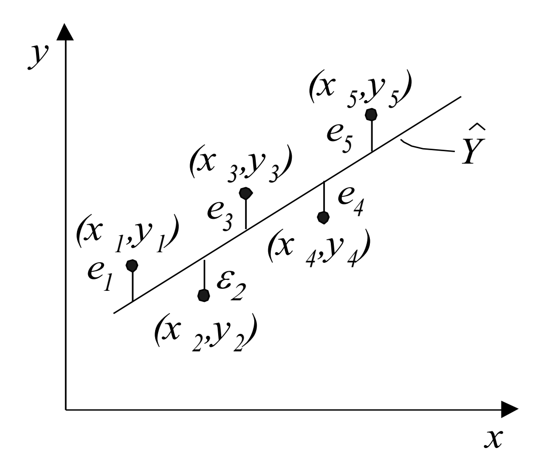


Рисунок 4.6 — График эксперимента

На рисунке 4.6 показаны точки (x_i, y_i), $i = \overline{1, N}$, полученные в машинном эксперименте с моделью системы. Делаем предположение, что модель результатов машинного эксперимента графически может быть представлена в виде прямой линии $\hat{Y} = b_0 + b_1 x$, где $\hat{Y}$ – величина, предсказываемая регрессионной моделью. Требуется

получить такое значение коэффициентов b_0 и b_1, при которых сумма квадратов ошибок является минимальной.

На рисунке 4.6 ошибка e_i для каждой экспериментальной точки определяется как расстояние по вертикали от этой точки до линии регрессии $\hat{Y}$.

$$e_i = \hat{Y}_i - Y_i = b_0 + b_1 x_i - Y_i,$$

$$\Phi = \sum_{i=1}^{N} \left(b_0 + b_1 x_i - Y_i\right)^2 \text{ — сумма квадратов ошибок,}$$

$$\left.\begin{aligned} \frac{\partial \Phi}{\partial b_0} &= 2\left(N b_0 + b_1 \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} Y_i\right) = 0 \\ \frac{\partial \Phi}{\partial b_1} &= 2 b_0 \sum_{i=1}^{N} x_i + 2 b_1 \sum_{i=1}^{N} x_i^2 - 2\sum_{i=1}^{N} x_i Y = 0 \end{aligned}\right\}$$

Решая систему нормальных уравнений получим

$$b_0 = \frac{\sum Y_i \sum x_i^2 - \sum x_i \sum x_i Y_i}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i^2\right)^2},$$

$$b_1 = \frac{N\sum x_i Y_i - \sum x_i \sum Y_i}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i^2\right)^2},$$

N — число реализаций при моделировании системы.

Мерой ошибки регрессионной модели служит среднеквадратическое отклонение

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N-2}} = \sqrt{\frac{\sum_{i=1}^{N} \left(b_0 + b_1 X_i - Y_i\right)^2}{N-2}},$$

Для нормального распределенных процессов приблизительно 67% точек находится в пределах одного отклонения σ_e от линии регрессии – труба *A*, и 95% точек – в пределах $2\sigma_e$ – труба *B* (рис. 4.7).

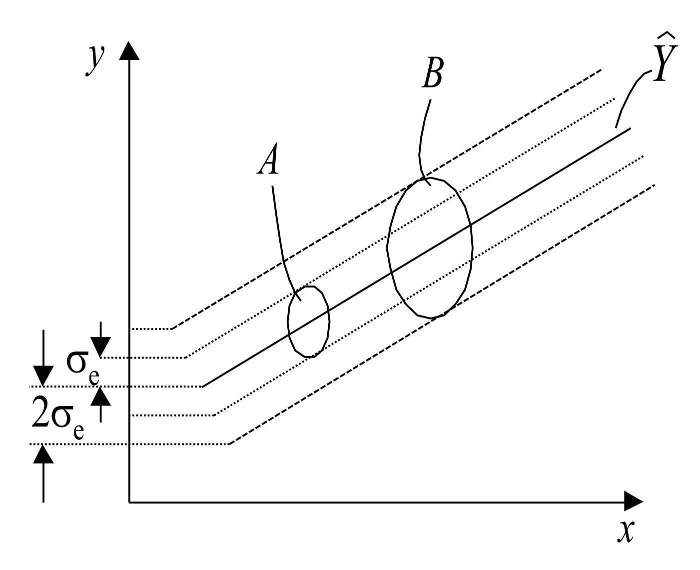


Рисунок 4.7 — Пределы отклонения экспериментальных точек

Для проверки точности оценок b_0, b_1 и адекватности модели используются критерии Стьюдента и Фишера.

Контрольные вопросы

1. Перечислите статистики генеральной совокупности и характеристики выборки.
2. Дайте определение функции распределения и постройте данную функцию для выборки.
3. Каким образом при экспериментировании соотносятся среднее выборки и математическое ожидание?
4. Каким образом при экспериментировании соотносятся среднее квадратическое отклонение и дисперсия?
5. Дайте определение корреляции, корреляционного момента и коэффициента корреляции.
6. Объясните понятие квантили распределения.
7. Объясните принцип получение интервальных оценок.
8. Каким образом получают оценку математического ожидания?
9. Каким образом получают оценку генеральной дисперсии?
10. Перечислите известные Вам методы получения псевдослучайных чисел. Подробно раскройте метод средних квадратов.
11. Подробно расскажите о получении последовательности случайных чисел мультипликативным конгруэнтным методом и смешанным методом.
12. В чем заключается проверка равномерности последовательности случайных чисел?

13. В чем заключается проверка независимости последовательности случайных чисел?

14. В чем заключается проверка стохастичности последовательности случайных чисел?

15. Каким образом проводят формирование возможных значений случайных величин с заданным законом распределения?

16. В чем заключается алгоритм получения случайных величин?

17. Какие правила автоматической остановки имитационного эксперимента Вы знаете?

5. Анализ результатов

5.1 Оценка результатов наблюдений

Целью статистической оценки является отыскивание оценки генерального параметра — идеального значения на основе выборочного параметра — экспериментального значения. Будем предлагать, что наблюдаемая случайная величина имеет нормальное распределение.

Основным оцениваемым параметром является генеральное среднее, т.е. математическое ожидание.

5.1.1 Оценка математического ожидания

Математическое ожидание легко определяется, если известна генеральная дисперсия σ^2. Зная σ^2, можно дать оценку для математического ожидания даже по одному наблюдению.

Пример 1. Определить доверительный интервал математического ожидания "а" по одному наблюдению $x = 3$, если известно, что генеральная совокупность имеет нормальное распределение с значением дисперсии $\sigma = 0.9$. В качестве доверительной вероятности возьмем P=0.96.

Берем в качестве доверительных границ симметричные квантили

$$u_{\frac{1-p}{2}} \le a \le u_{\frac{1+p}{2}}, \text{ или } u_{\frac{1-0,96}{2}} \le a \le u_{\frac{1+0,96}{2}},$$

Тогда имеем: $u_{0.02} \le a \le u_{0.98}$; по таблице $-2.06 \le a \le 2.06$.

Для общего нормального распределения из (4.5)

$$3 + \sigma \cdot u_{0.02} \le a \le 3 + \sigma \cdot u_{0.98}$$

$$3 + 0.9(-2.06) \le a \le 3 + 0.9 \cdot 2.06$$

$$1.146 \le a \le 4.854$$

Если над случайной величиной x проведено несколько наблюдений, то для оценки математического ожидания можно использовать выборочное среднее $\bar{x}$ и тогда $\bar{x} + \sigma_{\bar{x}} \cdot u_{\frac{1-p}{2}} \le a \le \bar{x} + \sigma_{\bar{x}} \cdot u_{\frac{1+p}{2}}$,

Если генеральная дисперсия известна, то $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, и окончательно получаем: $\bar{x} + \frac{\sigma}{\sqrt{n}} \cdot u_{\frac{1-p}{2}} \le a \le \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot u_{\frac{1+p}{2}}$,

Отсюда видно, что уменьшение интервальной оценки обратно пропорционально корню квадратному из числа наблюдений.

Пример 2. Оценить математическое ожидание "а" по генеральной дисперсии $\sigma^2 = 0.16$ и трем измерениям:

$x_1 = 7.2; x_2 = 7.8; x_3 = 7.6; n = 3.$

Возьмем уровень значимости $\alpha = 0.05$. Определим $\bar{x} = \dfrac{7.2 + 7.8 + 7.6}{3} = 7.53$; $\sigma = \sqrt{0{,}16} = 0{,}4$; $u_{1-\alpha/2} = 1.96$, т.к. вероятность p определена через уровень значимости $p = 1 - \alpha$, и $u_{\frac{1+p}{2}} = u_{1-\alpha/2}$; $u_{\frac{1-p}{2}} = u_{\alpha/2}$ или имеем $u_{\alpha/2} = -u_{1-\alpha/2}$. Тогда

$$\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} \le a \le \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2};$$

$$7.53 - \frac{\sqrt{0.16}}{\sqrt{3}} \cdot 1.96 \le a \le 7.53 + \frac{\sqrt{0.16}}{\sqrt{3}} \cdot 1.96;$$

$$7.07 \le a \le 7.99.$$

В этом примере приняли случайную величину x с нормальным законом распределения с параметрами a и $\dfrac{\sigma}{\sqrt{n}}$. Благодаря этому квантиль $u = \dfrac{a - \bar{x}}{\sigma}\sqrt{n}$ имеет стандартное нормальное распределение и с вероятностью $p = 1 - \alpha$ и удовлетворяет неравенству $-u_{1-\alpha/2} \le u \le u_{1-\alpha/2}$.

Однако в расчетах вместо σ^2 берут, как правило, выборочную дисперсию S^2. Это означает, что вместо квантили u нужно рассматривать величину $t = \dfrac{a - \bar{x}}{S}\sqrt{n}$.

При больших $n \ge 50$ величина S мало отличается от σ и поэтому значения u совпадают с t.

При малых же объемах выборок различие между t и u может оказаться существенным и, кроме того, само распределение величины t не является нормальным.

Используя общие законы теории вероятности, можно вывести формулы, описывающие распределение величины t. Это распределение называется t-распределением или распределением Стьюдента. Это распределение зависит только от степени свободы f, по которым подсчитывается дисперсия S^2. Если объем выборки равен n, то $f = n - 1$.

Закон распределения Стьюдента с $n-1$ степенями свободы имеет плотность

$$\varphi_{n-1}(t)=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\cdot\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}},$$

где Γ - известная гамма - функция.

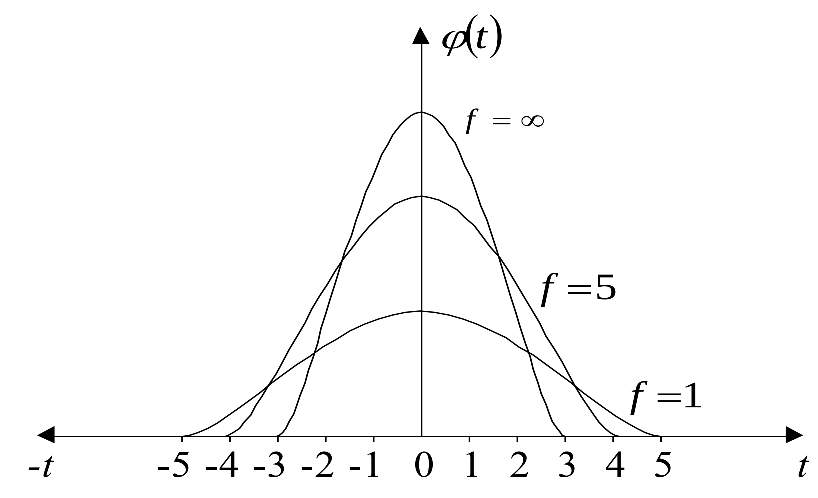


Рисунок 5.1 — График плотности t - распределения

Свойства плотности хорошо видны на графике плотности t — распределения (рис. 5.1). Они напоминают плотность нормального распределения. При $f\to\infty$ $(n\to\infty)$ выборочная дисперсия стремится к генеральной $(S^2\to\sigma^2)$ и распределение Стьюдента сближается с нормальным. При малых f распределение t сильно отличается от нормального. Поэтому роль t — распределения существенна в статистике малых выборок, еще называемой микростатистикой. Будем обозначать через t_α квантили t - распределения.

При доверительной вероятности $p=1-\alpha$ для величины t получаем доверительную оценку $-t_{1-\frac{\alpha}{2}}\le t\le t_{1-\frac{\alpha}{2}}$. Учитывая, что $t=\frac{a-\bar{x}}{S}\sqrt{n}$, получим $-t_{1-\frac{\alpha}{2}}\le\frac{a-\bar{x}}{S}\sqrt{n}\le t_{1-\frac{\alpha}{2}}$ или

$$\bar{x}-t_{1-\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}}\le a\le\bar{x}+t_{1-\frac{\alpha}{2}}\cdot\frac{S}{\sqrt{n}}$$

Эта оценка очень похожа на ранее рассматриваемые оценки, только здесь вместо σ берется S и поэтому вместо $u_{1-\frac{\alpha}{2}}$ необходимо брать $t_{1-\frac{\alpha}{2}}$.

Значения $t_{1-\frac{\alpha}{2}}$ для различных чисел степеней свободы f и уровней значимости α приводятся в соответствующих книгах по математической статистике.

Распределение Стьюдента позволяет оценить математическое ожидание при неизвестной генеральной дисперсии. При этом число наблюдений может быть очень малым.

5.1.2 Оценка генеральной дисперсии

Известно, что $M\left[S^2\right]=\sigma^2$. Поэтому, если известно распределение выборочной дисперсии S^2, то можно оценить и генеральную дисперсию σ^2.

Распределение величины S^2 можно получить с помощью распределения Пирсона, которое еще называют χ^2 — ***распределением*** (хи — квадрат распределение).

Для выборки с элементами $x_1, x_2 \ldots x_n$ через χ^2 обозначается сумма

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2 \tag{5.1}$$

Число степеней свободы здесь также равно $f = n - 1$(одно ограничение $\bar{x}$). *Плотность* χ^2*-распределения* зависит только от числа степеней свободы f (рис. 5.2).

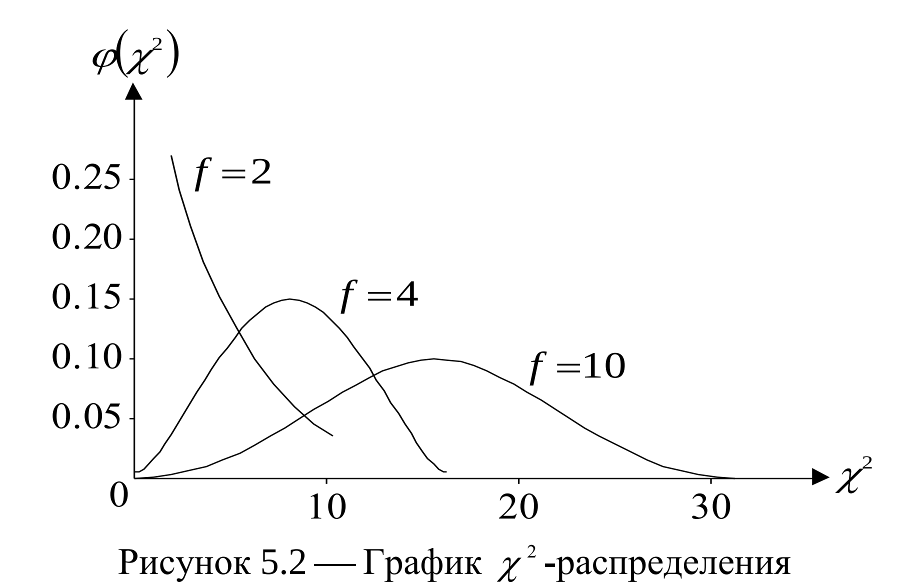


Рисунок 5.2 — График χ^2-распределения

Поскольку $\chi^2 \geq 0$, то и плотность расположена на промежутке $[0, \infty]$.

Кривые асимметричны, но степень асимметрии уменьшается

при увеличении f.

При уровне значимости α доверительная оценка величины χ^2 имеет вид $\chi^2_{\frac{\alpha}{2}} \le \chi^2 \le \chi^2_{1-\frac{\alpha}{2}}$.

Можно усмотреть связь между χ^2 и S^2. С учетом (5.1) можно записать $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{\sigma^2}{f}\chi^2$, откуда $\chi^2 = \frac{f\,S^2}{\sigma^2}$.

Отсюда с вероятностью $p = 1-\alpha$ справедливо неравенство

$$\chi^2_{\frac{\alpha}{2}} \le \frac{f\,S^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}} \text{ или } \frac{f\,S^2}{\chi^2_{1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{f\,S^2}{\chi^2_{\frac{\alpha}{2}}}.$$

Квантили $\chi^2_{\frac{\alpha}{2}}$ и $\chi^2_{1-\frac{\alpha}{2}}$ даются в таблицах.

5.2 Оценка параметров модели

Представим модели в следующем виде:

$$\eta = \eta(x, \beta),$$

где x — факторы (входные величины); β — неизвестные параметры (коэффициенты); η — реакция системы (выходная величина).

Целью анализа экспериментальных данных является определение оценок неизвестных параметров β в некоторой заданной области факторного пространства X. Рассмотрим статистическую модель (рис. 5.3).

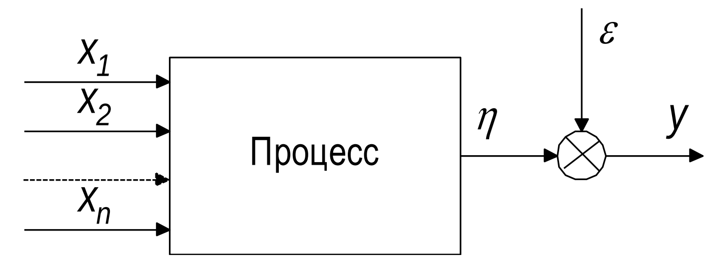


Рисунок 5.3 — Статистическая модель

В реальных условиях, из-за наличия помехи ε, экспериментатор измеряет величину y вместо истинного значения выходной величины η. Следовательно, опираясь на результаты измерения, нельзя получить абсолютно точных значений β. Вместо истинных параметров β приходится использовать случайные величины. Обозначим эти величины b и назовем их ***оценками*** β. Тогда, оцениваемое уравнение для модели будет иметь вид:

$$Y=Y(x,b) \qquad (5.2)$$

5.2.1 Общие требования, предъявляемые к оценкам параметров

Чтобы правильно и точно оценить параметры модели, оценки должны быть: *несмещенными, состоятельными, эффективными* и *достаточными*.

Оценки b являются ***несмещенными***, если их математические ожидания равны истинным значениям параметров:

$$M[b]=\beta.$$

Это значит, что в процессе вычисления параметров модели не должны возникать статистические ошибки.

Оценка называется ***состоятельной***, если при увеличении числа наблюдений n до бесконечности она сходится по вероятности к истинному параметру:

$$\lim_{n\to\infty} P[(b-\beta)\geq\varepsilon]=0, \text{ при любом } \varepsilon>0.$$

Достаточное условие для этого

$$\lim_{n\to\infty} M[(b-\beta)^2]=0$$

Оценки будут ***эффективными***, если они позволяют получить максимальную информацию из наблюдений. Часто бывает, что из исследуемого параметра можно найти несколько состоятельных оценок. Чтобы выбрать одну из них сравнивают дисперсии всех оценок и по минимуму дисперсии получают оценку, которая и будет эффективной

$$D[b]\leq D[\tilde{b}],$$

где $D[b]$ - дисперсия оценки b, $D[\tilde{b}]$ - дисперсия любых других несмещенных оценок b.

Пример. Дано n наблюдений случайной величины X. Возникает вопрос, какую величину принять за оценку математического ожидания: среднее выборки или медиану. Известно, что величина X распределена по нормальному закону с дисперсиями: $D[\bar{x}]=\frac{1}{n}\sigma^2$ и $D[x_{\frac{1}{2}}]=\frac{\pi}{2n}\sigma^2$, где $\bar{x}$ — среднее выборки; $x_{\frac{1}{2}}$ — медиана; n — объём выборки; σ^2 — дисперсия генеральной совокупности X. Так как $D[\bar{x}]<D[x_{\frac{1}{2}}]$, то оценка средней выборки будет эффективной.

Критерии несмещенности и эффективности следует рассматривать одновременно. Может оказаться, что смещенная

оценка с меньшей дисперсией будет более предпочтительной, чем несмещенная оценка с большей дисперсией.

5.2.2 Методы оценивания параметров

Существует несколько различных методов оценивания параметров:

- максимального правдоподобия;
- моментов;
- оценивание по Байесу;
- наименьших квадратов.

Метод максимального правдоподобия базируется на использовании априорной информации, полученной из эксперимента. Получают выборку значений случайной величины $X(x_1,x_2,...,x_n)$. Рассматривают оцениваемые параметры β как случайные величины с некоторым законом распределения вероятности. Затем это распределение перестраивается таким образом, чтобы получить апостериорное распределение вероятности, плотность которого несет информацию о возможных значениях β на основе экспериментальных данных X. Этот метод приводит к эффективным и состоятельным оценкам, однако оценки могут быть смещенными.

Метод моментов является одним из наиболее старых методов. При его использовании вычисляются первые n моментов случайной величины, которые затем приравниваются выборочным моментам. После этого находят n значений оцениваемых параметров β.

Оценивание по Байесу как и метод максимального правдоподобия основывается на использовании априорной информации. Определяется плотность распределения вероятностей x, и на основе апостериорной информации принимается решение.

Метод наименьших квадратов (МНК) является самым распространенным методом при оценивании параметров модели. Поэтому рассмотрим его более подробно на примере линейной модели с одной независимой величиной.

Уравнение модели с одной независимой величиной имеет вид:

$$\eta = \beta_0 + \beta_1 x \qquad (5.3)$$

Оценкой уравнения (5.3) будет:

$$\hat{y} = b_0 + b_1 x \qquad (5.4)$$

Уравнение (5.3) на плоскости представляет теоретическую линию регрессии, а (5.4) эмпирическую линию регрессии (рис. 5.4). Коэффициенты b_0 и b_1 являются оценками истинных коэффициентов β_0 и β_1.

На рисунке 5.4 обозначены точки (y_{ij},x_i) - одно измерение; $(\bar{y}_i, x_i)$ — выборочное среднее наблюдение при x_i; $(\hat{y}_i, x_i)$ — предсказанное значение выходной величины y_i при x_i; (η_i, x_i) —

истинное значение выходной величины η_i при x_i. Для несмещенных оценок $\eta_i = M[\bar{y}_i|_{x_i}]$, т.е. η_i есть математическое ожидание $\bar{y}_i$ при x_i.

По результатам опыта вычисляются коэффициенты b_0 и b_1 . Если бы все экспериментальные точки оказались на теоретической линии регрессии, то $\hat{y}_i - \eta_i = 0$ или

$$y_i - (\beta_0 + \beta_1 x_i) = 0,\ i=1,2,...n \tag{5.5}$$

и тогда коэффициенты β_0 и β_1 могли быть определены решением системы уравнений (5.5).

Однако, в реальных условиях левая часть (5.5) отличается от нуля на величину ε_i

$$\overset{*}{y}_i - (\beta_0 + \beta_1 x_i) = \varepsilon_i \tag{5.6}$$

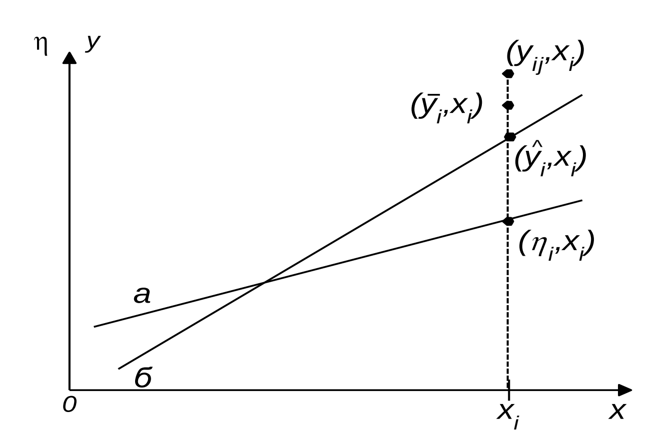


Рисунок 5.4 — Линии регрессии однофакторной модели
а — теоретическая линия регрессии; б — эмпирическая линия регрессии.

Величина ε_i называется ***невязкой***. Она может быть вызвана ошибкой эксперимента или неправильным выбором линейной модели. Поэтому возникает задача найти такие коэффициенты уравнения регрессии, при которых невязка будет минимальной. Лучшей оценкой является выражение $\sum_{i=1}^{n} \varepsilon_i^2 = \min$. Это выражение приводит к методу наименьших квадратов:

$$\Phi = \sum_{i=1}^{n} P_i(\bar{y}_i - \eta_i)^2 = \sum_{i=1}^{n} P_i(\bar{y}_i - \beta_0 - \beta_1 x_i)^2 \tag{5.7}$$

где P_i — число повторных измерений величины y при данном значении x_i.

Минимум функции Φ достигается при одновременном равенстве нулю частных производных этой функции по всем искомым коэффициентам:

$$
\begin{cases}
\dfrac{\partial \Phi}{\partial \beta_0} = \dfrac{\partial [\sum P_i(\bar{y}_i - \beta_0 - \beta_1 x_i)^2]}{\partial \beta_0} = -2\sum P_i(\bar{y}_i - \beta_0 - \beta_1 x_i) = 0 \\
\dfrac{\partial \Phi}{\partial \beta_1} = \dfrac{\partial [\sum P_i(\bar{y}_i - \beta_0 - \beta_1 x_i)^2]}{\partial \beta_1} = -2\sum P_i(\bar{y}_i - \beta_0 - \beta_1 x_i)x_i = 0
\end{cases} \quad (5.8)
$$

После замены β_0 и β_1 их оценками b_0 и b_1 получаем систему нормальных уравнений:

$$
\left.\begin{aligned}
&\sum P_i(\bar{y}_i - \beta_0 - \beta_1 x_i) = 0 \\
&\sum P_i(\bar{y}_i - \beta_0 - \beta_1 x_i)x_i = 0
\end{aligned}\right\} \quad \text{или}
$$

$$
\left.\begin{aligned}
&\sum P_i \bar{y}_i = \sum P_i b_0 + \sum P_i b_1 x_i \\
&\sum P_i \bar{y}_i x_i = \sum P_i b_0 x_i + \sum P_i b_1 x_i^2
\end{aligned}\right\} \quad (5.9)
$$

Решая систему нормальных уравнений относительно b_0 и b_1 получаем:

$$
\beta_0 \to b_0 = \frac{\sum P_i \bar{y}_i \sum P_i x_i^2 - \sum P_i x_i \sum P_i \bar{y}_i x_i}{\sum P_i \sum P_i x_i^2 - (\sum P_i x_i)^2} \quad (5.10)
$$

$$
\beta_1 \to b_1 = \frac{\sum P_i \sum P_i \bar{y}_i x_i - \sum P_i \sum P_i \bar{y}_i}{\sum P_i \sum P_i x_i^2 - (\sum P_i x_i)^2} \quad (5.11)
$$

5.3 Проблема обеспечения точности и достоверности результатов

Проблема обеспечения точности и достоверности результатов можно решить следующим образом. Обозначают точность оценки величиной E. Из-за наличия стохастичности ограниченности числа опытов N получают оценку $\tilde{E}$. Задаются неравенством

$$
\left| E - \tilde{E} \right| < \varepsilon \quad (5.12)
$$

и вероятностью P, что неравенство (5.12) выполняется

$$
Q = P\left\{ \left| E - \tilde{E} \right| < \varepsilon \right\}, \quad (5.13)
$$

где Q — называют достоверностью оценки. Вводят относительную точность оценки

$$
\varepsilon_0 = \frac{\varepsilon}{E}, \quad (5.14)
$$

с учетом которой достоверная оценка будет равна

$$
Q = P\left\{ \left| \frac{E - \tilde{E}}{E} \right| < \varepsilon_0 \right\}, \quad (5.15)
$$

Если известен закон распределения $\left| E - \tilde{E} \right|$, то с помощью анализа формулы (5.13) или (5.15) можно определить количество реализаций N.

В тех случаях, когда закон распределения $\left|E-\tilde{E}\right|$ найти не удается, то выдвигают предположение о характере закона распределения случайной величины E.

Рассмотрим взаимосвязь точности и достоверности, когда в качестве показателей эффективности E, выступают вероятность P, математическое ожидание a и дисперсия σ^2 .

Пусть вероятность появления некоторого события A, которое определяется состояниями процесса функционирования исследуемой системы, равна

$$P=P(A)$$

В качестве оценки вероятности P в данном случае выступает

$$\tilde{P}=\frac{m}{N},$$

где m - число положительных исходов. Тогда соотношение (5.13), связывающее точность и достоверность оценок с количеством реализаций, будет иметь вид

$$P\left\{\left|P-\frac{m}{N}\right|<\varepsilon\right\}=Q,$$

$$P\left\{P-\varepsilon<\frac{m}{N}<P+\varepsilon\right\}=Q \qquad (5.16)$$

Для ответа на вопрос о законе распределения величины $\tilde{P}$ представим эту частность в виде

$$\tilde{P}=\frac{m}{N}=\frac{1}{N}\sum_{i=1}^{N}x_i,$$

так как количество наступлений события A в данной реализации из N реализаций является случайной величиной ξ, принимающей значения $x_1=1$ с вероятностью P и $x_2=0$ с вероятностью $(1-P)$.

Математическое ожидание и дисперсия случайной величины ξ будут равны

$$M[\xi]=x_1P+x_2(1-P)=1\cdot P+0(1-P)=P,$$

$$D[\xi]=[x_1-M(\xi)]^2P+[x_2-M(\xi)]^2(1-P)=$$

$$=(1-P)^2P+(0-P)^2(1-P)=P(1-P)$$

Тогда

$$M\left[\tilde{P}\right]=M\left[\frac{m}{N}\right]=\frac{1}{N}M\left[\sum_{i=1}^{N}x_i\right]=\frac{1}{N}N\,M[\xi]=P$$

Это соотношение говорит о несмещенности оценки $\tilde{P}$ для вероятности P.

С учетом независимости значений величины x_i получаем

$$D\left[\tilde{P}\right] = D\left[\frac{m}{N}\right] = \frac{1}{N^2} D\left[\sum_{i=1}^{N} x_i\right] = \frac{1}{N^2} N\, D[\xi] =$$

$$= \frac{1}{N} P(1-P)$$

В силу центральной предельной теоремы теории вероятностей при достаточно больших N можно рассматривать частность $\frac{m}{N}$ как случайную величину, с нормальным законом распределения вероятностей с математическим ожиданием P и дисперсией $P(1-P)\cdot\frac{1}{N}$.

С учетом квантиля нормального распределения вероятностей t_φ точность оценки $\tilde{P}$

$$\varepsilon = t_\varphi \sqrt{\frac{1}{N} P(1-P)} \tag{5.17}$$

Количество реализаций для получения оценки $\tilde{P}$ с точностью ε и достоверностью Q будет равно

$$N = t_\varphi^2\, P(1-P)\cdot\frac{1}{\varepsilon^2} \tag{5.18}$$

Квантиль t_φ порядка $\varphi = \frac{1+Q}{2}$ находится из специальных таблиц.

Пример. Необходимо рассчитать количество реализаций N при статистическом моделировании системы S , когда в качестве показателя эффективности используется вероятность P при достоверности $Q = 0.95$ $(t_\varphi = 1.96)$ и точности $\varepsilon = 0.01; 0.02; 0.05$.

Ввиду того, что значения P до проведения эксперимента неизвестны, то вычисляют множество оценок N для диапазона возможных значений P, т.е. от 0 до 1 с дискретностью 0.1. Результаты расчетов с использованием выражения (5.18) представлены в таблице 5.1.

Из таблицы 5.1 видно, что при переходе от $P = 0.1$ (0.9) и $P = 0.5$ количество реализаций N возрастает примерно в 2,5 раза, а при переходе от $\varepsilon = 0.05$ и $\varepsilon = 0.01$ количество реализаций возрастает примерно в 25 раз.

Таблица 5.1

Результаты расчетов

Вероятность P	Точность ε

	0.05	0.02	0.01
0.1 (0.9)	140	900	3600
0.2 (0.8)	250	1500	6200
0.3 (0.7)	330	2100	8400
0.4 (0.6)	380	2300	9400
0.5 (0.5)	390	2400	9800

При тактическом планировании машинных экспериментов, когда значение *P* неизвестно, поступают следующим образом. Берут произвольно значение N_0, определяют по формуле $P_0 = \frac{m}{N_0}$, а затем по формуле (5.18) проводят вычисления, в которой вместо *P* подставляют P_0. Такая процедура оценки *N* может выполняться многократно.

Если отсутствует возможность получения каких-либо априорных сведений о вероятности *P*, то в таких случаях целесообразно задавать относительную точность результатов моделирования ε_0.

Для этих случаев формула (4.78) принимает вид

$$N = t_\varphi^2\, P(1-P)\frac{1}{P^2\varepsilon_0^2} = t_\varphi^2\,(1-P)\cdot\frac{1}{\varepsilon_0^2 P} \qquad (5.19)$$

Оценку показателя эффективности *E* модулируемой системы можно дать по результатам определения среднего значения некоторой случайной величины. Если случайная величина ξ имеет математическое ожидание *a* и дисперсию σ^2 и принимает в *i*-той реализации значение x_i, то в качестве оценки математического ожидания *a* принимают

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

При больших значениях *N* в силу предельной центральной теоремы теории вероятностей среднее арифметическое $\bar{x}$ будет иметь распределение, близкое к нормальному с математическим ожиданием *a* и дисперсией σ^2/N.

Точность оценки для математического ожидания равна $\varepsilon = t_\varphi \frac{\sigma}{\sqrt{N}}$, а количество реализаций

$$N = t_\varphi^2 \frac{\sigma^2}{\varepsilon_0^2 a^2} \qquad (5.20)$$

Если же в качестве показателя эффективности *E* выступает дисперсия σ^2, а в качестве ее оценки используется выборочная

дисперсия S^2,то математическое ожидание и дисперсия будут

$$M\left[S^2\right] = (N-1)\frac{\sigma^2}{N};$$

$$D\left[S^2\right] = \frac{\mu_4 - \sigma^4}{N},$$

где μ_4 — центральный момент четвертого порядка случайной величины.

Для частного случая, когда случайная величина имеет нормальное распределение, когда $\mu_4 = 3\sigma^4$, получаем

$$N = t_\varphi^2 \frac{2\sigma^4}{\varepsilon^2} = \frac{2t_\varphi^2}{\varepsilon_0^2} \qquad (5.21)$$

Для дисперсии σ^2 точность оценки

$$\varepsilon = t_\varphi \sqrt{\frac{\left(\mu_4 - \sigma^4\right)}{N}},$$

отсюда количество реализаций

$$N = t_\varphi^2 \left(\mu_4 - \sigma^4\right) / \varepsilon^2 \text{ или } N = t_\varphi^2 \left[\frac{\mu_4}{\sigma^4} - 1\right] \frac{1}{\varepsilon_0^2}.$$

Из выражений (5.19÷5.21) видно, что количество реализаций существенно зависит от дисперсии оцениваемой случайной величины. Поэтому выгодно выбирать те оцениваемые показатели эффективности E, которые имеют малые дисперсии.

5.4 Проблема уменьшения дисперсии оценок

Существуют методы, с помощью которых можно увеличить точность при заданном числе реализаций и, наоборот, при заданной точности оценок сократить число реализаций. Эти методы используют априорную информацию о структуре и поведении моделируемой системы.

В качестве такого метода можно использовать метод коррелированных реализаций (выборок), который используется в задачах сравнения двух или более альтернатив, например, можно сравнивать две системы S_1 и S_2.

Пусть для систем S_1 и S_2 оценки $\tilde{a}_1$и $\tilde{a}_2$ имеют дисперсии $D[\tilde{a}_1], D[\tilde{a}_2]$ и коэффициент корреляции оценок $\tilde{a}_1, \tilde{a}_2$ равен $R[\tilde{a}_1, \tilde{a}_2]$, то дисперсию погрешности оценки $\tilde{d} = \tilde{a}_1 - \tilde{a}_2$ разности $d = a_1 - a_2$ можно найти из соотношения

$$D\left[\tilde{d}\right] = D[\tilde{a}_1 - \tilde{a}_2] = D[\tilde{a}_1] + D[\tilde{a}_2] - 2R[\tilde{a}_1, \tilde{a}_2] \cdot \sigma_1 \sigma_2, \qquad (5.22)$$

где $\sigma_1 = \sqrt{D[\tilde{a}_1]}, \sigma_2 = \sqrt{D[\tilde{a}_2]}$.

При независимом моделировании вариантов системы с

использованием различных реализаций псевдослучайных последовательностей коэффициент корреляции оценок равен

$$R[\tilde{a}_1, \tilde{a}_2] = 0,$$

$$D_H[\tilde{d}] = D[\tilde{a}_1] + D[\tilde{a}_2] \quad (5.23)$$

следовательно, $D[\tilde{d}] < D_H[\tilde{d}]$,т.к. при имитационном моделировании удается получить $R > 0$.

5.5 Регрессионный анализ

После нахождения коэффициентов модели возникает задача установить пригодность модели и значимость коэффициентов. С этого момента метод наименьших квадратов превращается в регрессионный анализ. Применение регрессионного анализа возможно только при выполнении следующих предположений.

1. Математическое ожидание величины $\bar{y}_i$ при заданном значении x_i является линейной функцией по параметрам, т.е. модели должны быть линейными по параметрам.

2. Значения x_i не являются случайными величинами.

3. Дисперсия ошибки ε_i равна дисперсии величины $\bar{y}_i$: $\sigma^2_{\varepsilon_i} = \sigma^2_{\bar{y}_i}$. Эта дисперсия может быть постоянной либо зависимой от *x*.

4. Различные измерения величины *y* взаимно независимы.

При выполнении этих четырех условий МНК дает несмещенные оценки b_0 и b_1 параметров β_0 и β_1 .

В случае нахождения доверительной области для коэффициентов β_0 и β_1 должно выполняться еще одно предположение.

5. Условие распределения $\bar{y}_i$ при заданном значении x_i нормально относительно математического ожидания $\eta_i = M[\bar{y}_i|_{x_i}]$.

5.6 Проверка адекватности модели

Для того чтобы принять решение относительно модели необходимо проверить гипотезу: линейная модель по параметрам удовлетворительно описывает экспериментальные данные.

Проверку адекватности модели выполняют с помощью *критерия Фишера*:

$$F = \frac{S_r^2}{S_e^2},$$

где S_r^2 — дисперсия неточности модели, S_e^2 — дисперсия ошибки эксперимента.

Если $F < F_{1-\alpha}$, то гипотеза о том, что линейная модель адекватна, принимается. Здесь $F_{1-\alpha}$ — табличное значение; α —

уровень значимости; F — вычисленное значение.

При $F > F_{1-\alpha}$ гипотеза адекватности линейной модели по параметрам отвергается.

Обратимся к рисунку 5.4. Дисперсия S_e^2 характеризует ошибку эксперимента, которую можно получить путем возведения в квадрат разности $(y_{ij} - \bar{y}_i)$ с последующим суммированием по всем экспериментальным точкам:

$$S_e^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{P_i}(y_{ij} - \bar{y}_i)^2}{\sum_{i=1}^{n} P_i - n} \tag{5.24}$$

Знаменатель дисперсии равен разности между общим количеством экспериментальных точек $\sum_{i=1}^{n} P_i$ и числом наложенных связей n, т.к. каждая выборка дает одну связь.

Дисперсия S_r^2 характеризует неточность подгонки модели, её неадекватность. Поэтому её можно получить из разности $(\bar{y}_i - \hat{y}_i)$, которая после возведения в квадрат и суммирования представляет собой сумму квадратов отклонений относительно эмпирической линии регрессии:

$$S_r^2 = \frac{\sum_{i=1}^{n} P_i(\bar{y}_i - \hat{y}_i)^2}{n-2} \qquad (4.23)(5.25)$$

Здесь число степеней свободы равно *(n-2)*. Две связи вызваны двумя ограничениями, связанными с определением b_0 и b_1.

Если математическое ожидание дисперсии S_r^2 равно дисперсии выходной величины η:

$$M[S_r^2] = \sigma_\eta^2,$$

то полученная эмпирическая модель будет корректной.

5.7 Определение дисперсии воспроизводимости эксперимента

Дисперсия S_y^2 характеризует ошибку эксперимента и называют ***дисперсией воспроизводимости эксперимента***. Она может быть получена путем объединения дисперсий S_e^2 и S_r^2:

$$S_y^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{P_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{n} P_i(\bar{y}_i - \hat{y}_i)^2}{(\sum_{i=1}^{n} P_i - n) + (n-2)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{P_i}(y_{ij} - \hat{y}_i)^2}{\sum_{i=1}^{n} P_i - 2} \tag{5.26}$$

Более простое определение $S_y^2 = S_e^2$.

5.8 Проверка однородности дисперсий

Простейшим критерием проверки однородности дисперсий является критерий Фишера, который представляет из себя отношение большей дисперсии к меньшей:

$$F = \frac{S_{\max}^2}{S_{\min}^2}$$

Полученная величина F сравнивается с табличной величиной $F_{1-\alpha}$. Если $F > F_{1-\alpha}$, то дисперсии значимо отличаются друг от друга и тогда они неоднородны. Если сравнивается много дисперсий (более двух) и одна из дисперсий значимо превышает остальные, то можно применять критерий Кохрена – это отношение максимальной дисперсии к сумме всех дисперсий:

$$G = \frac{S_{\max}^2}{\sum_{i=1}^{n} S_i^2},$$

Гипотеза об однородности подтверждается, если $G < G_t$, где G_t — табличное значение.

5.9 Принятие решений после построения модели процесса

Остановимся, в начале, на принятии решения, когда линейная модель адекватна.

При этом возможны три варианта.

1. Все коэффициенты регрессии значимы.
2. Часть коэффициентов регрессии значима, а часть незначима.
3. Все коэффициенты регрессии незначимы.

Для первого варианта могут быть приняты следующие решения: окончания исследования, переход к планам второго порядка и движение по градиенту. Переход ко второй серии планирования для модели второго порядка позволяет получить экстремум параметра.

Во втором варианте приходится выдвигать гипотезу, что влияние данных эффектов незначимо. Причиной незначимости коэффициентов может быть неудачный выбор интервалов варьирования и включение второстепенных факторов, большая ошибка опыта и т.д.

Для этого либо расширяют интервалы варьирования по факторам, либо увеличивают число параллельных опытов.

И последний вариант, линейная модель адекватна, все коэффициенты не значимы. Это может произойти вследствие большой ошибки эксперимента или узких интервалов варьирования.

Контрольные вопросы

1. В чем заключается оценка математического ожидания?
2. В чем заключается оценка генеральной дисперсии?
3. В чем заключается оценка параметров модели?
4. Перечислите общие требования, предъявляемые к оценкам параметров.
5. Какие методы оценивания параметров Вы знаете?
6. В чем заключается проблема обеспечения точности и достоверности результатов?
7. В чем заключается проблема уменьшения дисперсии оценок?
8. Каким образом проводят проверку адекватности модели?
9. Как определилить дисперсию воспроизводимости эксперимента?
10. В чем заключается проверка однородности дисперсий ?
11. Какие решения могут быть приняты после построения модели процесса?

6 Математические схемы описания технических систем

6.1 Непрерывно–детерминированные модели (D — схемы)

Математические схемы такого вида отражают *динамику* процессов, протекающих во времени в системе. Поэтому они называются ***D — схемы.*** Частным случаем динамических систем являются *системы автоматического управления*.

Линейная автоматическая система описывается линейным дифференциальным уравнением вида

$$a_0 \frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \ldots + a_n y = b_0 \frac{d^m x}{dt^m} + b_1 \frac{d^{m-1} x}{dt^{m-1}} + \ldots + b_m x \qquad (6.1)$$

где *x(t)* — задающее воздействие или входная переменная системы; *y(t)* — состояние системы или выходная переменная; a_i, b_i — коэффициенты; *t* — время.

На рисунке 6.1 представлена укрупненная функциональная схема системы автоматического управления, где $h'(t) = x(t) - y(t)$ — сигнал ошибки; $h''(t)$ — управляющее воздействие; *f(t)* — возмущающее воздействие. Данная система основана на принципе отрицательной обратной связи, так как для приведения выходной переменной *y(t)* к ее заданному значению используется информация об отклонении между ними. По ней можно разработать структурную схему и математическую модель в виде передаточной функции или в виде дифференциального уравнения (6.1), в котором для простоты предполагается, что точки приложения возмущающих воздействий совпадают с входом системы.

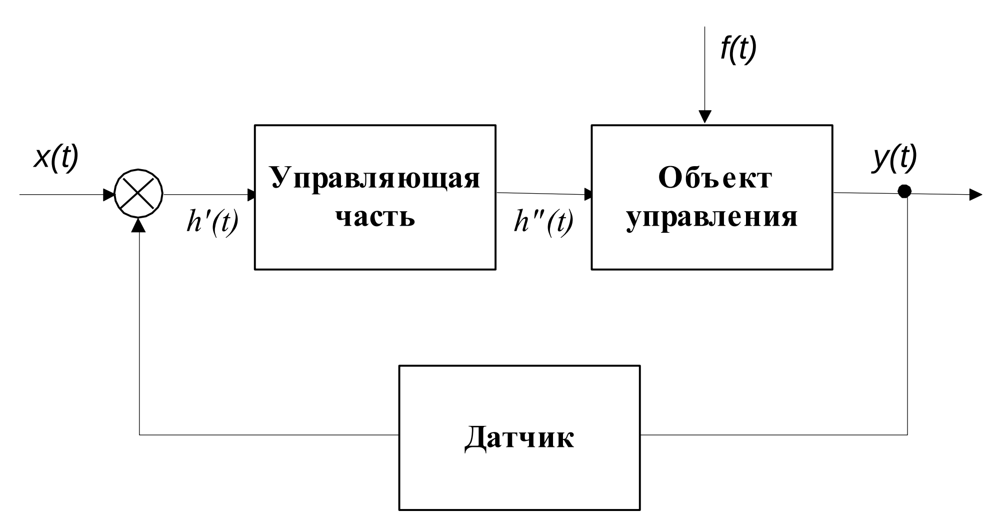


Рисунок 6.1 — Структура системы автоматического управления

Непрерывно — детерминированные схемы (D — схемы) выполняются на аналоговых вычислительных машинах (АВМ).

6.2 Дискретно–детерминированные модели (F — схемы)

Основным видом дискретно–детерминированных моделей является *конечный автомат.*

Конечным автоматом называют дискретный преобразователь информации, способный под воздействием входных сигналов переходить из одного состояния в другое и формировать сигналы на выходе. Это автомат *с памятью*. Для организации памяти в описание автомата вводят автоматное время и понятие *состояние автомата*.

Понятие «*состояние»* автомата означает, что выходной сигнал автомата зависит не только от входных сигналов в данный момент времени, но и учитывает входные сигналы, поступающие ранее. Это позволяет устранить время как явную переменную и выразить выходные сигналы как функцию состояний и входных сигналов.

Всякий переход автомата из одного состояния в другое возможен не ранее, чем через дискретный интервал времени. Причем сам переход считается, происходит мгновенно, то есть не учитывают переходные процессы в реальных схемах.

Существует два способа введения автоматного времени, по которому автоматы делятся на *синхронные* и *асинхронные*.

В *синхронных* автоматах моменты времени, в которых фиксируются изменения состояний автомата, задаются специальным устройством — генератором синхросигналов. Причем сигналы поступают через равные интервалы времени — Δt. Частота тактового генератора выбирается такой, чтобы любой элемент автомата успел закончить свою работу до появления очередного импульса.

В *асинхронном* автомате моменты перехода автомата из одного состояния в другое заранее не определены и зависят от конкретных событий. В таких автоматах интервал дискретности является переменным.

Также существуют *детерминированные* и *вероятностные* автоматы.

В *детерминированных* автоматах поведение и структура автомата в каждый момент времени однозначно определяются текущей входной информацией и состоянием автомата.

В *вероятностных* автоматах они зависят от случайного выбора.

Абстрактно конечный автомат можно представить как математическую схему (F — схему), которая характеризуется шестью видами переменных и функций:

1) конечное множество $x(t)$ входных сигналов (входной алфавит);
2) конечное множество $y(t)$ выходных сигналов (выходной алфавит);
3) конечное множество $z(t)$ внутренних состояний (алфавит

состояний);

4) начальное состояние автомата z_0 , $z_0 \in Z$;

5) функция переходов $\varphi(z, x)$ автомата из одного состояния в другое;

6) функция выходов $\Psi(z, x)$ автомата.

Абстрактный конечный автомат имеет один вход и один выход. В каждый дискретный момент времени $t=0,1,2,...$ F — автомат находится в определенном состоянии $z(t)$ из множества Z — состояний автомата, причем в начальный момент времени $t=0$ он всегда находится в начальном состоянии $z(0)=z_0$. В момент t , будучи в состоянии $z(t)$, автомат способен воспринять на входном канале сигнал $x(t) \in X$ и выдать на выходном канале сигнал $y(t) = \Psi[z(t), x(t)]$, переходя в состояние

$$z(t+1) = \varphi[z(t), x(t)] \text{ , где } z(t) \in Z, x(t) \in X .$$

Абстрактный конечный автомат реализует некоторое отображение множества слов входного алфавита X на множество слов выходного алфавита Y , то есть, если на вход конечного автомата, установленного в начальное состояние z_0 , подавать в некоторой последовательности буквы входного алфавита $x(0), x(1), x(2),...$, которые составляют входное слово, то на выходе автомата будут последовательно появляться буквы выходного алфавита $y(0), y(1), y(2),...$ образуя выходное слово.

Следовательно, работа конечного автомата происходит по следующей схеме: на каждом t — ом такте на вход автомата, находящегося в состоянии $z(t)$, подается некоторый сигнал $x(t)$, на который автомат реагирует переходом на $(t+1)$– ом такте в новое состояние $z(t+1)$ и выдачей некоторого выходного сигнала.

В зависимости от способа определения выходного сигнала синхронные абстрактные конечные автоматы подразделяются на два типа:

– F — автомат первого рода, также называется ***автомат Мили***:

$$\begin{cases} z(t+1) = \varphi[z(t), x(t)], t = 01,2,...; \\ y(t) = \Psi[z(t), x(t)], t = 0,1,2,...; \end{cases}$$

– F — автомат второго рода:

$$\begin{cases} z(t+1) = \varphi[z(t), x(t)], t = 01,2,...; \\ y(t) = \Psi[z(t), x(t-1)], t = 1,2,3,...; \end{cases}$$

Автомат второго рода, для которого

$$y(t) = \Psi[z(t)], t = 0,1,2,...$$

называется ***автомат Мура*** — функция выходов не зависит от входной переменной $x(t)$.

Чтобы задать конечный F — автомат, необходимо описать все элементы множества $F = \langle z, x, y, \varphi, \Psi, z_0 \rangle$.

Существует несколько способов задания работы F — автоматов среди которых наибольшее применение нашли табличный, графический и матричный.

6.3 Дискретно — непрерывные модели

Процессы в линейных импульсных и цифровых системах автоматического управления описываются дискретно — разностными уравнениями вида:

$$b_k \Delta^k y(n) + b_{k-1} \Delta^{k-1} y(n) + \ldots + b_0 y(n) = x(n), \qquad (6.2)$$

где *x(n)* — решетчатая функция входного сигнала; *y(n)* — решетчатая функция выходного сигнала, которая определяется решением уравнения (6.2); b_k — постоянные коэффициенты; $\Delta^k y(n)$ – разность *к* – го порядка; *t=nT*, где *nT* – *n*–ый момент времени, *T* — период дискретности (в выражении (6.2) период условно принят за единицу).

Уравнение (6.2) можно представить в другом виде:

$$a_k y(n-k) + a_{k-1} y(n-k+1) + \ldots + a_0 y(n) = x(n) \qquad (6.3)$$

Уравнение (6.3) представляет собой рекуррентное соотношение, которое позволяет вычислить любой *(i+1)*–й член последовательности по значениям предыдущих её членов *i,i-1,...* и значению *x(i+1).*

Основным математическим аппаратом моделирования цифровых автоматических систем является Z — преобразование, которое базируется на дискретном преобразовании Лапласа. Для этого необходимо найти импульсную передаточную функцию системы, задаться входной переменной и, варьируя параметрами системы, можно найти лучший вариант проектируемой системы.

6.4 Дискретно — стохастические модели (Р — схемы)

К дискретно — стохастической модели относится *вероятностный автомат*. В общем, виде вероятностный автомат является дискретным потактным преобразователем информации с памятью, функционирование которого в каждом такте зависит только от состояния памяти в нем и может быть описано статистически. Поведение автомата зависит от случайного выбора.

Применение схем вероятностных автоматов имеет важное значение для проектирования дискретных систем, в которых проявляется статистически закономерное случайное поведение.

Для Р — автомата вводится аналогичное математическое понятие, как и для F — автомата. Рассмотрим множество G, элементами которого являются всевозможные пары (x_i, z_s), где x_i и z_s

элементы входного подмножества X и подмножества состояний Z соответственно. Если существуют две такие функции φ и Ψ, что с их помощью осуществляется отображение $G \to Z$ и $G \to Y$, то говорят, что $F = \langle Z, X, Y, \varphi, \Psi \rangle$ определяет автомат детерминированного типа.

Функция переходов вероятностного автомата определяет не одно конкретное состояние, а распределение вероятностей на множестве состояний (автомат со случайными переходами). Функция выходов также есть распределение вероятностей на множестве выходных сигналов (автомат со случайными выходами).

Для описания вероятностного автомата введем в рассмотрение более общую математическую схему. Пусть Ф — множество всевозможных пар вида (z_k, y_j), где y_j — элемент выходного подмножества Y. Далее потребуем, чтобы любой элемент множества G индуцировал на множестве Ф некоторый закон распределения следующего вида:

элементы из Ф ... $(z_1, y_1)(z_1, y_2)...(z_k, y_{j-1})(z_k, y_j)$

(x_i, z_s) ... b_{11} b_{12} ... $b_{k(j-1)}$ b_{kj}

где b_{kj} — вероятности перехода автомата в состояние z_k и появления на выходе сигнала y_j, если он был в состоянии z_s и на его вход в этот момент времени поступал сигнал x_i.

Число таких распределений, представленных в виде таблиц равно числу элементов множества G. Если обозначить это множество таблиц через B, то тогда четверку элементов $P = \langle Z, X, Y, B \rangle$ называют ***вероятностным автоматом*** (P — автоматом). При этом $\sum_{k=1}^{k} \sum_{j=1}^{j} b_{kj} = 1$.

Частным случаем P — автомата, задаваемого как $P = \langle Z, X, Y, B \rangle$ являются автоматы, у которых либо переход в новое состояние, либо выходной сигнал определяются детерминировано(Z — *детерминированный вероятностный автомат, Y — детерминированный вероятностный автомат* соответственно).

Очевидно, что с точки зрения математического аппарата задание Y — детерминированного P — автомата эквивалентно заданию некоторой марковской цепи с конечным множеством состояний. В связи с этим аппарат марковских цепей является основным при использовании P– схем для аналитических расчетов. Подобные P — автоматы используют генераторы марковских последовательностей при построении процессов функционирования систем или воздействий внешней среды.

Марковские последовательности, согласно теореме Маркова

[9], — это последовательность случайных величин, для которой справедливо выражение

$$\lim_{N\to\infty}\left(\frac{1}{N^2}\right)D\left[\sum_{i=1}^{N}x_i\right]=0,$$

где N – количество независимых испытаний; D — дисперсия.

Такие P — автоматы (P — схемы) могут быть использованы для оценки различных характеристик исследуемых систем как для аналитических моделей, так и для имитационных моделей с использованием методов статистического моделирования.

Y — детерминированный P — автомат можно задать двумя таблицами: переходов (табл.6.1) и выходов (табл.6.2).

Таблица 6.1

Таблица переходов

z_k	z_k				
	z_1	z_2	...	z_{k-1}	z_k
z_1	P_{11}	P_{12}	...	$P_{1(k-1)}$	P_{1k}
z_2	P_{21}	P_{22}	...	$P_{2(k-1)}$	P_{2k}
...	...	...	...	...	...
z_k	P_{k1}	P_{k2}	...	$P_{k(k-1)}$	P_{kk}

Таблица 6.2

Таблица выходов

Z...	z_1	z_2	...	z_{k-1}	z_k
Y...	y_{i1}	y_{i2}	...	$y_{i(k-1)}$	y_{ik}

Где P_{ij} — вероятность перехода P — автомата из состояния z_i в состояние z_j, при этом $\sum_{j=1}^{k}P_{ij}=1$.

Таблицу 6.1 можно представить в виде квадратной матрицы размерности $k\times k$. Такую таблицу будем называть *матрицей переходных вероятностей* или просто *матрицей переходов P — автомата*, которую можно представить в компактной форме:

$$P_p=\begin{vmatrix}P_{11} & P_{12} & \dots & P_{1k}\\ P_{21} & P_{22} & \dots & P_{2k}\\ . & . & . & .\\ P_{k1} & P_{k2} & \dots & P_{kk}\end{vmatrix}$$

Для описания Y — детерминированного P — автомата необходимо задать начальное распределение вероятностей вида:

Z...	z_1	z_2	...	z_{k-1}	z_k
D...	d_1	d_2	...	d_{k-1}	d_k

где d_k вероятность того, что в начале работы P — автомат находится в состоянии z_k, при этом $\sum_{k=1}^{K} d_k = 1$.

И так, до начала работы P — автомат находится в состоянии z_0 и в начальный (нулевой) такт времени меняет состояние в соответствии с распределением D. После этого смена состояний автомата определяется матрицей переходов P. С учетом z_0 размерность матрицы P_p следует увеличить до $(k+1)\times(k+1)$, при этом первая строка матрицы будет $(d_0,d_1,d_2,...,d_k)$, а первый столбец будет нулевым.

Пример. Y– детерминированный P– автомат задан матрицей переходов:

$$P_p = \begin{vmatrix} 0 & 0{,}50 & 0 & 0 & 0{,}50 \\ 0 & 0 & 0 & 1{,}00 & 0 \\ 0 & 0 & 0{,}75 & 0 & 0{,}25 \\ 0 & 0 & 0{,}40 & 0 & 0{,}60 \\ 0 & 1{,}00 & 0 & 0 & 0 \end{vmatrix}$$

и таблицей выходов

Таблица 6.3

Таблица выходов

Z	z_0	z_1	z_2	z_3	z_4
Y	0	0	1	1	0

С учетом матрицы переходов граф переходов вероятностного автомата представлен на рисунке 6.2.

Требуется оценить суммарные финальные вероятности пребывания этого автомата в состоянии z_2 и z_3, т.е. когда на выходе автомата появляются единицы.

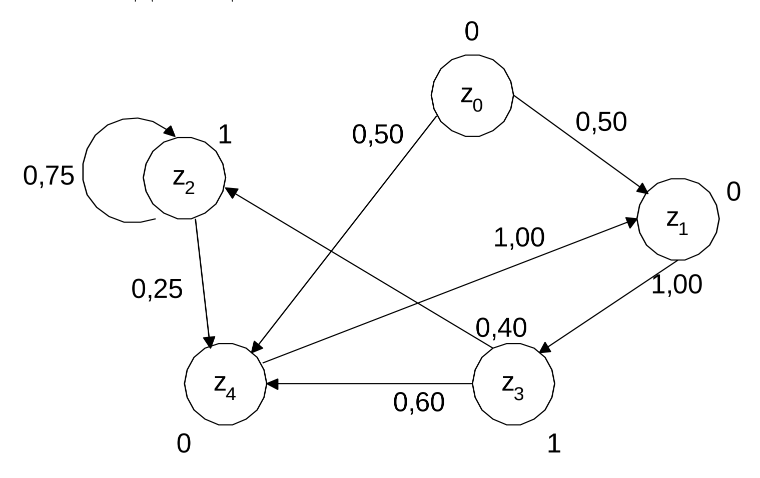


Рисунок 6.2. — Граф переходов

При аналитическом подходе можно использовать известные соотношения из теории марковских цепей и получить систему уравнений для определения финальных вероятностей. Причем начальное состояние можно не учитывать в виду того, что начальное распределение не оказывает влияние на значения финальных вероятностей. Тогда матрица переходов без первого столбца и строки примет вид:

$$\vec{C} = \vec{C}\begin{vmatrix} 0 & 0 & 1{,}00 & 0 \\ 0 & 0{,}75 & 0 & 0{,}25 \\ 0 & 0{,}40 & 0 & 0{,}60 \\ 1{,}00 & 0 & 0 & 0 \end{vmatrix};$$

где $\vec{C} = (c_k) = (c_1, c_2, c_3, c_4)$ — финальная вероятность пребывания Y — детерминированного Р — автомата в состоянии z_k.

В результате получаем систему уравнений:

$$\begin{cases} c_1 = c_4 \\ c_2 = 0{,}75c_2 + 0{,}40c_3 \\ c_3 = c_1 \\ c_4 = 0{,}25c_2 + 0{,}60c_3 \end{cases} \qquad (6.4)$$

К данной системе следует добавить условие нормировки:

$$c_1 + c_2 + c_3 + c_4 = 1 \qquad (6.5)$$

Теперь решая систему уравнений (6.4) совместно с (6.5), получаем:

$$c_1 = \frac{5}{23}; c_2 = \frac{8}{23;} c_3 = \frac{5}{23}; c_4 = \frac{5}{23}.$$

Таким образом, при бесконечной работе заданного автомата на его выходе будет формироваться двоичная последовательность с вероятностью появления единицы, равной: $c_2 + c_3 = \frac{13}{23} = 0{,}5652$.

Кроме аналитических моделей в виде Р — схем можно применять и имитационные модели, реализуемые, например, методом статистического моделирования.

6.5 Непрерывно–стохастические модели (Q — схемы)

Такие модели рассмотрим на примере использования в качестве типовых математических схем систем массового обслуживания, которые называют ***Q — схемами***. Такие Q — схемы применяются при формализации процессов функционирования систем, которые по своей сути являются процессами *обслуживания* [5, 9,10].

К *процессам обслуживания* можно отнести: потоки поставок продукции некоторому предприятию, потоки деталей и комплектующих изделий на сборочном конвейере цеха, заявки на обработку информации ЭВМ от удаленных терминалов сети ЭВМ. Характерным признаком для функционирования таких систем или сетей является случайное появление заявок на обслуживание. Причем в любом элементарном акте обслуживания можно выделить две основные составляющие: ожидание обслуживания и , собственно, сам процесс обслуживания заявки. Представим это в виде некоторого i-го прибора обслуживания $П_i$ (рис.6.3), состоящего из накопителя заявок $Н_i$, в котором может находится одновременно λ_i заявок; $К_i$ – канал обслуживания заявок.

На каждый элемент прибора $П_i$ поступают потоки событий, в накопитель $Н_i$ поток заявок λ_i, на канал $К_i$ – поток обслуживания $И_i$.

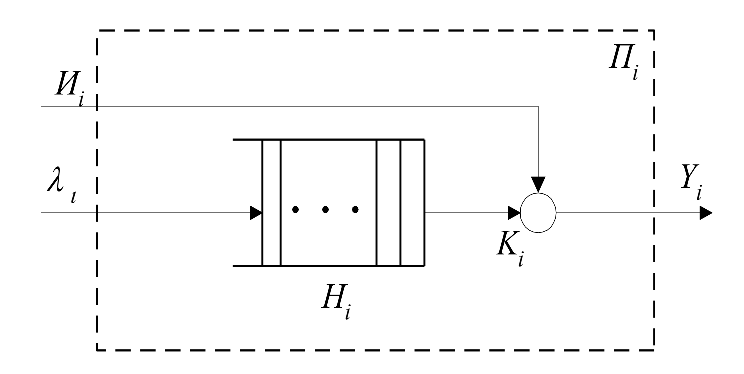


Рисунок 6.3 — Прибор обслуживания

Потоки событий могут быть *однородными*, если он характеризуется только последовательностью поступления этих событий ($0 \le t_1 \le t_2 \le ... \le t_n$), или *неоднородными*, если он характеризуется набором признаков события, например таким набором признаков: источник заявок, наличие приоритета, возможность обслуживания тем или иным типом канала и т.п.

Обычно при моделировании различных систем применительно к каналу $К_i$ можно считать, что поток заявок λ_i на входе $К_i$ образует подмножество неуправляемых переменных, а поток обслуживания $И_i$ – образует подмножество управляемых переменных.

Те заявки, которые по различным причинам не обслуживаются каналом $К_i$, образуют выходной поток $У_i$.

Приборы могут подключаться последовательно и (или) параллельно, что предоставляет возможность моделировать процессы массового обслуживания для разных объектов. Такой принцип

называется *агрегатированием.*

6.6 Стохастические минимаксные модели

Эти модели можно отнести к оптимальным стохастическим моделям.

Во многих случаях при построении модели не все условия заранее известны. Эффективность нахождения модели здесь будет зависеть от трех факторов:

- заданных условий $x_1, x_2,...,x_n$;
- неизвестных условий $y_1,y_2,...,y_k$;
- зависящих от нас факторов $u_1,u_2,...,u_m$, которые необходимо найти.

Показатель эффективности решения такой задачи имеет вид:

$$W = W(x_1, x_2,...,x_n; y_1, y_2,...,y_k; u_1, u_2,...,u_m)$$

Наличие неизвестных условий y_i переводит задачу оптимизации в задачу о выборе решения в условиях неопределенности. Задача становится чрезвычайно сложной.

Особенно задача осложняется для случаев, когда величины y_i не обладают статистической устойчивостью, то есть неизвестные условия y_i нельзя изучить с помощью статистических методов. Их законы распределения либо не могут быть получены, либо вовсе не существуют.

В этих случаях рассматриваются комбинации всевозможных значений Y: $\alpha = (y_1^*, y_2^*,..., y_k^*)$ таким образом, чтобы получить как «наилучшее», так и «наихудшее» сочетания значений переменных y_i.

Тогда в качестве критерия оптимизации рассматривается:

$$W = W(x, \alpha, u)$$

$$x = x(x_1, x_2,...,x_n), \alpha = \alpha(\alpha_1, \alpha_2,...,\alpha_k)$$

$$u = u(u_1, u_2,...,u_m)$$

Критерий Вальда

При использовании этого критерия всегда ориентируются на то, что случайные факторы сложатся в наихудшую комбинацию. Такой критерий называют ***пессимистическим***:

$$\max_u \min_\alpha W(x, \alpha, u).$$

Пример. Имеется четыре стратегии игрока: u_1,u_2,u_3,u_4; три варианта условий: y_1,y_2,y_3 и известные значения x_{ij} i=1,2,3,4, j=1,2,3.

Требуется найти оптимальное решение с использованием критерия Вальда.

Составим матрицу игры наихудших результатов (табл. 6.4). Согласно критерия Вальда: $\alpha_i = \alpha_3 = 0{,}25$.

Таблица 6.4

Матрица игры

$u_i \setminus y_i$	y_1	y_2	y_3	α_i
u_1	0,20	0,30	0,15	0,15
u_2	0,75	0,20	0,35	0,20
u_3	0,25	0,80	0,25	0,25
u_4	0,85	0,05	0,45	0,05

Следовательно оптимальной будет стратегия u_3.

Критерий Гурвица

При этом критерии предполагается, что с вероятностью β сложатся наилучшая, а с вероятностью (1-β) наихудшая комбинации случайных факторов:

$$\max_u [\beta \max_\alpha W(x,\alpha,u) + (1-\beta) \min_\alpha W(x,\alpha,u)] \tag{6.6}$$

$$0 < \beta < 1$$

При β=0 критерий Гурвица переходит в критерий Вальда, при β=1 получается так называемый критерий «здорового оптимизма». Коэффициент β выбирается из субъективных соображений. Если ситуация очень опасная, то его выбирают ближе к 1.

Пример. Берется тот же пример (табл.6.4). Принимается β=0,4. Достраивается таблица 6.4 столбцом, соответствующим оптимальной оценке (табл. 6.5). Максимальное значение η_3=0,47 соответствует стратегии u_3.

Таблица 6.5.

Матрица игры

$u_i \setminus y_i$	y_1	y_2	y_3	α_i	ω_i	η_i
u_1	0,20	0,30	0,15	0,15	0,30	0,21
u_2	0,75	0,20	0,35	0,20	0,75	0,42
u_3	0,25	0,80	0,25	0,25	0,80	0,47
u_4	0,85	0,05	0,45	0,05	0,85	0,37

По критерию (6.6) получаем: $\eta_i = 0{,}4\omega_i + 0{,}6\alpha_i$.

Контрольные вопросы

1. Для каких объектов применяются непрерывно — детерминированные модели (D — схемы)? Перечислите особенности этой схемы.
2. Для каких объектов применяются дискретно —

детерминированные модели (F — схемы)? Перечислите особенности этой схемы.

3. Для каких объектов применяются дискретно — непрерывные модели? Перечислите особенности этой схемы.

4. Для каких объектов применяются дискретно — стохастические модели (P — схемы)? Перечислите особенности этой схемы.

5. Для каких объектов применяются непрерывно — стохастические модели (Q — схемы)? Перечислите особенности этой схемы.

6. Для каких объектов применяются стохастические минимаксные модели? Перечислите особенности этой схемы.

Литература

1. Авдеев О.Н., Мотайленко Л.В. Моделирование систем: Учебное пособие. СПб.: Изд-во СПбГТУ, 2001. 179 с.
2. Автоматизация проектирования вычислительных систем. Языки моделирования и базы данных: Пер. с англ./ Под ред. Бейера. М: Мир, 1979.
3. Бусленко Н.П. Моделирование сложных систем. М.: Наука, 1978.
4. Власов Л.В., Колесников Д.Н., Сорокин И.А. Имитационное моделирование систем массового обслуживания с использованием GPSS: Учеб. пособие. Л.: ЛПИ, 1989.
5. Дурандин К.П. и др. Моделирование сложных систем с использованием сетей массового обслуживания: Учеб. пособие. Л.: ПИ (СПбГТУ), 1981.
6. Колесников Д.Н., Сиднев А.Г., Юрганов А.А. Моделирование случайных факторов в задачах автоматики и вычислительной техники: Учеб. пособие. СПб.: Изд. СПбГТУ, 1994.
7. Кузин Л.Т. Основы кибернетики. М.: Энергия, 2 тома: том 1 «Математические основы кибернетики», 1973г.; том 2 «Основы кибернетических моделей». 1979.
8. Максимей И.В. Имитационное моделирование на ЭВМ. М.: Радио и связь,1988.
9. Марков А.А. Моделирование информационно-вычислительных процессов: Учебное пособие для вузов. М.: Изд-во МГТУ им. Н.Э. Баумана, 1999. 360с.
10. Петерсон Дж. Теория сетей Петри и моделирование систем. М: Мир,1984.
11. Советов Б.Я., Яковлев С.А. Моделирование систем: Учеб. для вузов — 2-е изд., перераб. и доп. М.: Высш. шк., 2001. 343 с.
12. Строгалев В.П., Толкачева И.О. Имитационное моделирование: Учебное пособие. М.: Изд-во МГТУ им. Н.Э.Баумана, 2008. 280 с.
13. Тюрин Ю.Н., Марков А.А. Анализ данных на компьютере: Пер. с англ./ Под ред. В.Э. Фигурнова М.: Финансы и статистика, 1995.
14. Шеннон Р. Имитационное моделирование систем – искусство и наука / Пер. с англ. под ред. Е.К. Масловского. М.: Изд. "Мир", 1978. 418с.

Учебное издание

Лилия Владимировна Мотайленко

ОСНОВЫ МОДЕЛИРОВАНИЯ

Учебное пособие
для студентов всех форм обучения направления подготовки 230400
«Информационные системы и технологии»

Компьютерная вёрстка: Л.В. Мотайленко
Технический редактор: Л.В. Мотайленко
Подписано в печать 25.12.2013. Формат 60×90/16.
Гарнитура Times New Roman. Усл. п. л.
Тираж 100 экз. Заказ №

Адрес издательства:
Россия, 180000, г. Псков, ул. Л. Толстого, 4,
Издательство ПсковГУ